

Bridging Nonliving and Living Matter

Abstract Assembling non-biological materials (geomaterials) into a proto-organism constitutes a bridge between nonliving and living matter. In this article we present a simple step-by-step route to assemble a proto-organism. Many pictures have been proposed to describe this transition within the origins-of-life and artificial life communities, and more recently alternative pictures have been emerging from advances in nanoscience and biotechnology. The proposed proto-organism lends itself to both traditions and defines a new picture based on a simple idea: Given a set of required functionalities, minimize the physicochemical structures that support these functionalities, and make sure that all structures self-assemble and mutually enhance each other's existence. The result is the first concrete, rational design of a simple physicochemical system that integrates the key functionalities in a thermodynamically favorable manner as a lipid aggregate integrates proto-genes and a proto-metabolism. Under external pumping of free energy, the metabolic processes produce the required building blocks, and only specific gene sequences enhance the metabolic kinetics sufficiently for the whole system to survive. We propose an experimental implementation of the proto-organism with a discussion of our experimental results, together with relevant results produced by other experimental groups, and we specify what is still missing experimentally. Identifying the missing steps is just as important as providing the road map for the transition. We derive the kinetic and thermodynamic conditions of each of the proto-organism subsystems together with relevant theoretical and computational results about these subsystems. We present and discuss detailed 3D simulations of the lipid aggregation processes. From the reaction kinetics we derive analytical aggregate size distributions, and derive key properties of the metabolic efficiency and stability. Thermodynamics and kinetics of the ligation directed self-replication of the proto-genes is discussed, and we summarize the full life cycle of the proto-organism by comparing size, replication time, and energy with the biomass efficiency of contemporary unicells. Finally, we also compare our proto-organism picture with existing origins-of-life and protocell pictures.

By assembling one possible bridge between nonliving and living matter we hope to provide a piece in the ancient puzzle about who we are and where we come from.

Steen Rasmussen^{1,2}

Liaohai Chen³

Martin Nilsson¹

Shigeaki Abe^{1,3}

¹Self-Organizing Systems

EES-6, MS-T003

Los Alamos National Laboratory

Los Alamos, NM 87545

steen@lanl.gov

nilsson@lanl.gov

²Santa Fe Institute

1399 Hyde Park Road

Santa Fe, NM 87506

³Bioscience Division

Argonne National Laboratory

Argonne, IL 60439

lhchen@anl.gov

sabe@anl.gov

Keywords

Proto-cells, origins of life, nanotechnology, supermolecular chemistry, self-assembly, physical chemistry, photochemistry, multiscale simulation, evolution, artificial chemistry

I Introduction

Like contemporary living cells, under appropriate laboratory conditions future proto-organisms will sustain themselves chemically, feed from the environment, self-reproduce, be capable of evolution, and be able to die under environmental stress. Unlike actual cells and their proto-cell ancestor, engineered proto-organisms will have an artificial origin, environment, and metabolism. They will not be made or derived from existing cells; instead they will be built “from scratch” to operate only in artificial environments. Although they may have some metabolic pathways common to those in natural cells, their designed metabolisms may exhibit completely different chemistries. Proto-organisms have not yet been constructed in the laboratory, partly due to the physicochemical complexity of assembling such structures and partly because only very few focused efforts exist worldwide. However, recently several efforts have started.

There are two significantly different approaches to synthesizing primitive life forms: a *bottom-up* and a *top-down* approach. The current contribution is based on a bottom-up approach concerned with constructing a simple living system from nonliving organic and inorganic materials through self-assembly, with metabolic processes driven by an external supply of free energy. The top-down approach concerns itself with systematic simplification of very simple existing cells.

The presented work on assembling a proto-organism has grown out of a bottom-up tradition, which has emerged over the past ten years at the intersection between the origins-of-life studies, nanoscience, new materials, supermolecular chemistry, and artificial life activities. Our work addresses several of the open questions formulated by the artificial life community at the 2000 conference in Portland, which can be found in Bedau et al. [4]. The work mainly addresses question 1, “Generate a molecular proto-organism in vitro,” but it also touches question 4, “Simulate a unicellular organism over its entire lifecycle,” and question 5, “Explain how rules and symbols are generated from physical dynamics in living systems.”

I.1 Bottom Up

Until recently it was difficult to separate the ideas on how to bridge nonliving and living matter from the theories of the possible origins of life. Experimental bridging pathways that were believed to be likely have often been proposed as particular origins-of-life pictures, and these pictures have often been tightly linked to specific molecules and/or processes. Pictures less dependent on the physicochemical details usually lack experimental support, but may still be valuable from a theoretical point of view.

The *naked-gene* approach to bridging nonliving and living matter has for more than a generation dominated the origins-of-life debate, which from the mid-eighties [12] through the nineties proposed that the RNA was probably “the first living molecule.” The *RNA world* [29, 28] supporters had a very compelling argument: since RNA has both catalytic and information storage capabilities, it could act simultaneously as DNA and protein do in contemporary life. Many different polymerization and replication approaches have been developed [24]. However, to develop a self-replicating RNA molecule turned out to be a very difficult task [39], and this problem has recently taken some of the attractiveness from the RNA world picture.

The *peptide world*, stating that proteins probably were the first biomolecules, perhaps constitutes the oldest physicochemically based picture of the origins of life [68]. It is based on the fact that proteins can polymerize from amino acids in a prebiotic environment. The later experiments by Miller and Urey from 1959 [57] demonstrated laboratory production of amino acids as well as many other key biological compounds

under harsh prebiotic-like conditions. Today we know that most violent or extreme processes (under high pressure, temperature, or irradiation) that involve simple elements produce traces of rather complex biomolecules. Examples of such processes include high explosive blasts [32] and cosmic chemistry in meteorites [16]. One of the latest and most intriguing experimental developments within the peptide world is the generation of a self-reproducing peptide system by Ghadiri and coworkers [45, 82].

An alternative origins-of-life picture, the *lipid world*, was originally developed by Luisi and coworkers [47] and by Morowitz, Deamer and others in 1988 [60], who have later worked towards assembling proto-cells based on a self-reproducing lipid vesicle encapsulating a self-replicating RNA proto-gene. Self-reproducing lipid aggregates (micelles and liposomes, single and bilayer lipid structures) have been developed successfully in several laboratories. It is also known that lipids are produced by cosmic chemistry and probably were readily available at the time of the origins of life. An interesting new twist on the lipid world came recently with the proposed role of atmospheric aerosols in the origins of life [94]. For example, encapsulating of self-assembling structures, such as microtubules, within liposomes has also served in attempts to develop partial proto-cell models [35].

Recently, important players from the RNA world community have joined this encapsulation approach, as discussed by Szostack et al. [89]. Their vision consists of a bottom-up construction, but with rather high-level building blocks. They propose using an RNA-based chemistry, very similar to the information chemistries in contemporary natural cells. As they acknowledge, the complexity of the catalysis needed to reproduce RNA (using an RNA replicase that is still a long way from being able to replicate its own coding RNA) and the need for metabolic encoding of the building blocks (lipids and RNA) remain barriers to their approach. In addition, there is no natural physicochemical integration between the RNA and the lipid container.

Another proto-cell vision has been developed by Pohorille and Deamer [70]. The starting point for their proposal is also the self-assembling dynamics of lipids to form lipid bilayer vesicles (liposomes); however, they suggest encapsulating modern cell organelles. They discuss different versions of artificial cells for different biotechnological purposes, with variations on RNA- and DNA-based information chemistries, as well as metabolisms, noting, as do Szostak et al. [89], that there are significant technical hurdles to making these complex chemistries work even when using existing cellular organelles. They conclude:

Many individual components needed for such structures have already been developed and many others are likely to be constructed in the near future. The main challenge now is to encapsulate them in a single cellular compartment and ensure that they will work in concert in a controlled manner.

This critique to some extent applies to all bottom-up approaches.

The bottom-up approach also has attracted many theoretical and computational contributions. The notion of a hypercycle is a powerful theoretical concept proposed by Eigen in 1971 [20]. The hypercycle is a cooperative structure of self-replicating proto-genes. Also in 1971, Ganti [27] proposed a cooperative structure coupling a container, a metabolism, and a genetic system, which at least at a conceptual level is quite similar to our proposed proto-organism. This notion of cooperative structures being a key element for the bootstrapping of chemical systems to becoming biological systems has been the most important driver for the theory-driven origin-of-life approaches. Such approaches span rather realistic model systems with autocatalytic sets of polymers as developed by Farmer et al. in 1986 [23], Kauffman in 1986 [42], and Bagley and Farmer in 1991 [3]. Lately, an interesting new idea has

emerged based on autocatalytic lipid structures, due to Segre et al. [83]. Wachterhauser [98] has proposed mineral-surface-based metabolic processes as the key bootstrap for the first self-reproducing lipid systems. Random graph generalizations of the cooperative feedback concept for proto-genetic systems were developed by Rasmussen in 1985 [75] and 1988 [76], and spatial hypercycle generalizations were developed by Hogeweg and coworkers [6]. Study of spatially extended systems showed that the hypercycle organization persists in the face of parasites in such settings, due to the formation of spiral waves. As a major simplification it was shown by McCaskill and coworkers [54, 55] that hypercycles, or other reaction based cyclic cooperative feedback structures, are not necessary for the stabilization of distributed catalysis when one uses proximity in space. Recently, abstract self-reproducing computational proto-cells have been developed by Ikegami and coworkers [67]. These proto-cells have more realistic physical properties encoded than (for example) the intriguing self-reproducing loops proposed by Langton in 1984 [44] and by Sayama in 1998 [84], and for that reason they are more informative for the construction of physicochemical proto-cells.

In addition to the notion of cooperative structures in chemical networks, two other theoretical and computational traditions are important for our proto-organism work: (a) the detailed molecular dynamics (MD) simulation approach [100, 51] together with (b) the traditional lattice gas [7], lattice Boltzmann [41], and Ginzburg-Landau [38] approaches. Adding a technique for intermediate time- and length-scale simulation, the MD lattice gas [52, 66], we have the necessary components for a developing a predictive, 3D, virtual proto-organism simulation. Rajagopalan [74] has compiled a nice review of self-assembly simulations, but a comprehensive multiscale simulation is not yet available.

Our proto-organism design is inspired both by the above theoretical and computational ideas and by the experimental concepts for a proto-cell due to Luisi and coworkers [49], but our model also differs from this and other proposals in several aspects:

1. Our focal point is a minimalistic, thermodynamic coupling between the three functional structures: container, metabolism, and genes. We do not start with a self-replicating container or a self-replicating gene, which is then combined.
2. Instead of RNA, we envision simpler molecules, such as peptide nucleic acids (PNAs) [64, 63]. Due to their hydrophobic backbone, these may be easier to couple with the lipid layer than RNA, and they are also easier to synthesize.
3. Instead of sophisticated ribozymes, we consider very simple short oligos that are capable of enzymeless self-replication by means of a ligation mechanism [96, 97].
4. As in the other proto-cell proposals, we utilize the lipid to keep the cooperative structure together, although the proto-genetic activity is not on the inside of a vesicle [1] but on the outside of a lipid aggregate. We can therefore work with simpler lipid structures such as micelles.
5. We make extensive use of the differences between the thermodynamic properties of the lipid phase and those of the water and the lipid-water interface, and as a result we obtain a quite different chemistry.

1.2 Top Down

The top-down approach starts with contemporary cells with very small genomes. Experiments with a simple bacterium, *Mycoplasma genitalium* [26, 36] indicate that cells can have much of their genome removed and remain alive. Current estimates of the minimal

genome size based on this approach are ≈ 300 –350 genes. The minimal genome size obtained from these experiments may be considered an upper bound on the genome size expected for simple proto-cells based on translated protein chemistry. Readjustment to deletions via retro-evolution of the cells and complementation has not been fully investigated. These minimal cells may still have artefacts in their genome that are needed to support physical and metabolic structures that have been required along the evolutionary path of the original bacteria, but that may not be needed for proto-cells. Recently Venter and Smith [30] have received financial support from the U.S. government (DOE) to create new, more primitive life forms using this top-down approach.

In parallel with this experimental top-down approach, a push for whole-cell simulations is under way within the systems biology community; see for example *Science*, 299 (2002). These efforts are driven by a growing understanding of the details of the metabolic pathways and genetic networks. This tradition is concerned with representation and simulation of life as it is, as opposed to the artificial life community's focus on a synthesis and simulation of life as it could be. It is worth noting, however, that these whole-cell simulations could in a natural manner test and benchmark their approaches by simulating the much simpler proto-cells currently under development—including the proposed proto-organism in this article.

1.3 Structure of this Article

This article consists of several parts. Some constitute a traditional reporting of work done, with associated results. Others deal with work that is not yet done, but that is of key importance for the whole proto-organism project. If any one of these few missing links turns out to be unachievable, our design will have to be changed. However, due to the general interest in the problem of generating life, we feel it is still of value to present this new picture. The mixture of tested and non-tested processes has, however, made it challenging to write this article. Secondly, the article has been challenging to write because it seeks to synthesize concepts and methods from very different traditions. We have attempted to address the latter problem by a sandwich construction with a general beginning and a general summary, interleaved with the necessary technical details.

Section 2 is a detailed conceptual presentation of the proto-organism. Section 3 discusses the experimental implementations, both what we know and what we don't know, and is rather technical. It requires some prior chemical knowledge and some familiarity with experimental techniques and procedures. Section 4 is also rather technical, as it discusses the theoretical underpinnings of Section 3 and thus requires some prior knowledge of physical chemistry and computational physics. Unfortunately, some technical jargon is unavoidable to limit the length of these two sections. Although this construction allows us to be precise and to the point as we present the different technical issues both in the experimental and in the theoretical and computational context, its presentation has the reader revisit the key technical issues from three different perspectives: conceptual, experimental, and theoretical. However, we are attempting to address a diverse and interdisciplinary community, so we apologize if the text seems too repetitive for the domain experts. Section 4.5 (on the full life cycle of the proto-organism) is a summary of key elements from Sections 3 and 4, and it is again written in a more conceptual manner, as are the discussion and the conclusion (Sections 5 and 6).

2 Organizational and Functional Structure of a Proto-organism

Which observable functional properties should the proto-organism have? An extensive discussion exists in the literature on the topic “What is life?” which we shall not enter here. For our purpose, it is sufficient to note that the ability (a) to evolve, (b) to self-reproduce, (c) to metabolize, (d) to respond adaptively to environmental changes,

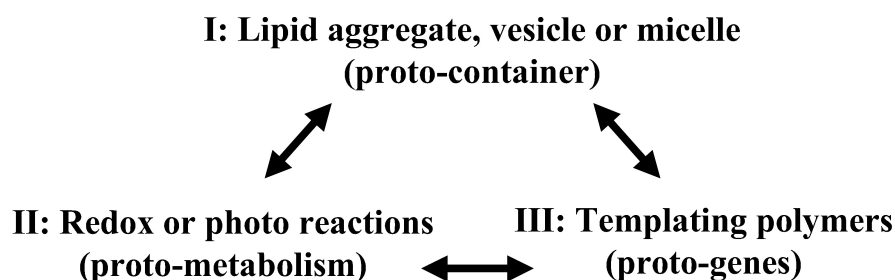


Figure 1. The functional organization of the aggregates involved in the simple proto-organism. Multiple physicochemical implementations of this structure are possible, but they all pose difficult thermodynamic and kinetic challenges.

and (e) to die are normally referred to as key properties of a living system. The proto-organism we are proposing has all of the above properties, and they are generated as the components of the proto-organism assemble and use free energy to digest appropriate precursor molecules. To address the question of how to implement these five functionalities in a minimalist manner, we may ask: Which molecular aggregate interactions can carry or generate these functionalities?—or, in other words—How can we assemble a dynamical hierarchy that ensures that these functionalities are generated? An extensive discussion of why we believe it is useful to operate with dynamical hierarchies is found in Rasmussen et al. [77].

We are now ready to present a thermodynamically downhill, step-by-step process that combines into a single cooperative aggregate a proto-container, a proto-metabolism, and proto-genes constituting a proto-organism. This physicochemical network derives energy from a coupled redox complex or photochemical reactions—a simple form of metabolism—and carries encoded information about the metabolic processes in a proto-gene, which together with the metabolic complexes are integrated in a lipid aggregate. Thus, this aggregate can self-replicate, use energy and nutrients available from its environment, undergo evolutionary change over time, and ultimately die. The high level organization of the proto-organism is shown in Figure 1.

A set of possible physicochemical implementations of the processes involved will now be presented together with a discussion of the thermodynamic and cooperative principles. A detailed discussion of the chemistry, thermodynamics, and kinetics follows in the next sections.

The simplest way to assemble a proto-organism seems to be through the combination of two physicochemical systems: a lipid-metabolic (redox or photochemical) system and a lipid-templating system. In Figure 2 we discuss the structure and function of these systems and how they can be integrated. We seek to avoid initially getting lost in the physicochemical details so that the main conceptual design can be communicated as clearly as possible, although we still want to include enough of the necessary details to clarify the main thermodynamic and kinetic issues.

The top part of Figure 2 (I) describes the self-assembly of the lipid container. The left-hand side (II) describes the steps involved in the integration of a simple metabolic system in an amphiphilic aggregate (micelle or vesicle bilayer). The right-hand side (III) describes the integration of a templating polymer in an amphiphilic aggregate (micelle or vesicle bilayer). Finally, when these three systems are integrated (I+II+III), they constitute a simple proto-organism. Note that in this example, with redox-driven metabolic processes, we assume that the lipid aggregate is a vesicle.

As shown in (1) in Figure 2 at the top (I), amphiphiles self-assemble into micelles and vesicles (or liposomes) that can host organic metabolic molecules. Since many

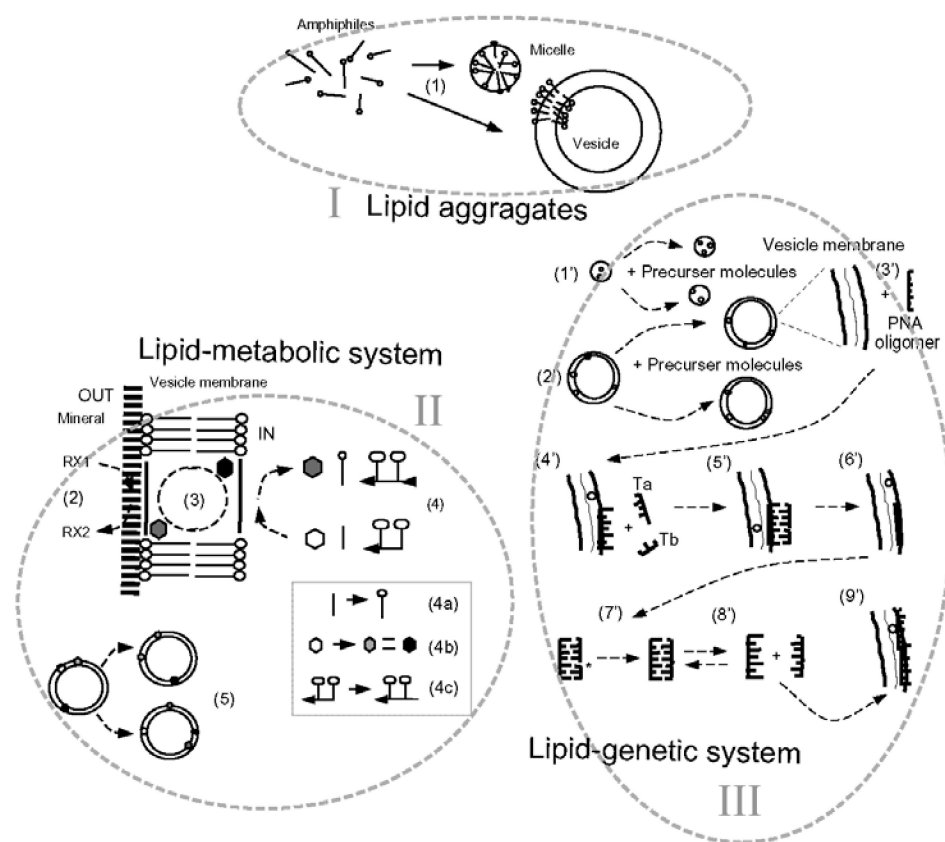


Figure 2. Two experimental systems II and III need to be combined to form a simple proto-organism. The basis for II and III is the self-assembly of lipid aggregates driven by the hydrophobic effect as shown in I. The left-hand side II comprises the redox (or photochemical) reactions within the lipid aggregate, and the right-hand side III comprises the templating reactions within the lipid aggregate. Either (chemical) redox or photochemical energy is driving the proto-metabolic processes for subsystem II. The proto-genetic system, subsystem III, is based on a delicate balance among the hydrophobic effect, hydrogen bond formation and dissociation, and a ligation reaction. See text for details.

organic redox or photoactive molecules are hydrophobic, thermodynamics will drive them to become integrated into the hydrophobic portion of the lipids. The energy source used by this metabolic system may be either chemical energy or light. As an example of the utilization of chemical energy, (system II) in Figure 2 shows a vesicle that contains organic redox molecules, which becomes attached to a mineral surface containing appropriate metal atoms, due to charge differences between the head groups in the aggregate and the mineral interface.

Assuming that the hydrophobic redox molecules in the lipid phase and metal atoms in the mineral surface have appropriate redox potentials, they can act as electron donors and receptors in an energy-rich exchange. As shown in (2) and (3), the redox molecules receive or donate electrons at the mineral surface and move to the inside of the lipid phase, where the electrons can be exchanged. This induces processes (4) that, for example, generate more amphiphilic (4a) and/or organic redox molecules (4b) and/or parts of templating polymers (4c), from appropriate precursor molecules.

Vesicles self-reproduce (5) as more amphiphiles are produced, increasing a vesicle's surface area until the structure becomes unstable and surface tension breaks it into two new vesicles. The more interface area between vesicles and mineral surfaces,

the more electron transfer reactions, which in turn will generate more amphiphiles and organic redox molecules. The more redox molecules there are, the more redox molecules and the more lipid molecules can be produced. Thus, through two coupled autocatalytic reactions, cooperation can emerge between self-reproducing vesicles and a simple production of redox molecules.

The right-hand side picks up where the left-hand side ended, with the self-reproduction of micelles (1') and/or vesicles (2'). Development of a proto-gene begins on the outer surface of a vesicle membrane (3'). The templating polymer is an oligomer with a hydrophobic backbone that sinks into the vesicle's surface (4').¹ As an example, this polymer may be PNA or a related substance, because it has a simple backbone like that of a hydrophobic polypeptide rather than the more-complicated and charged RNA and DNA backbones composed of phosphates and sugars.² Since PNA has the same nucleobases as DNA or RNA, however, templating is possible. Note that without a hydrophobic backbone it would not be thermodynamically favorable for the templating polymer to sink into the lipid aggregate. For simplicity we use PNA throughout this paper as a representative templating polymer with a hydrophobic backbone. Note that the attachment of (for example) RNA on a charged lipid interface would be possible, but the following reaction steps would be difficult.

The next step is to have two shorter, complementary polymer pieces, labeled "Ta" and "Tb," recognize the hydrogen-bond sites on the information template and form a double-stranded molecule (5'). Once this thermodynamic recognition has occurred, the templated molecule sinks further into the interior of the lipid aggregate (6') so that ligation can occur (at the asterisk) within the hydrophobic region (7'). Here a polymerization is more thermodynamically favorable,³ and the ligation could be driven by extraction of water (according to le Chatelier's principle). In addition the ligation process could be enhanced in several ways—for example, with a simple catalyst or in connection with a drying cycle of the lipid. Again, we want to emphasize that ligation is highly thermodynamically unfavorable in water; also note that the charged backbones of RNA or DNA would not allow these molecules to sink into the hydrophobic phase.

The kinetics and thermodynamics of the association-dissociation processes for the two now complementary templating polymers within the vesicle must be in balance for dissociation to occur (8'). If the processes are balanced, both templating polymers will be able to perform template-directed self-replication and the process can be repeated (9'). The result of this process would be a template-directed replication of the proto-gene polymer.

To obtain full cooperation between our vesicles, organic redox molecules, and proto-genes, the strands in the vesicle must, at a minimum, enhance one or more of the redox processes. Specific strands (e.g., conformation or chemical properties of nucleobases) could (for example) enhance the rate of formation of either the amphiphilic or the organic redox molecules, at which point specific functionality would be encoded as a given sequence of monomers within the templating polymer. Such a feedback loop between the proto-genes, the redox processes, and the lipid aggregate enables a Darwinian evolution: In the simplest situation, assume we have two different proto-gene templates in two different aggregates: one that encodes a more efficient set of redox reactions, and one that has no positive influence on the redox reactions. Assuming everything else equal, a faster growth of the cooperative gene aggregate would result, and if there were competition for resources (available energy and building

1 The idea of thermodynamically coupling the template and the lipid aggregate in this manner was initially developed by Klaus Lackner and Steen Rasmussen in 1997.

2 A detailed discussion of the properties of the hydrophobic PNA backbone is given in Sections 3 and 4.

3 The ligation process is discussed in detail in Sections 3 and 4.

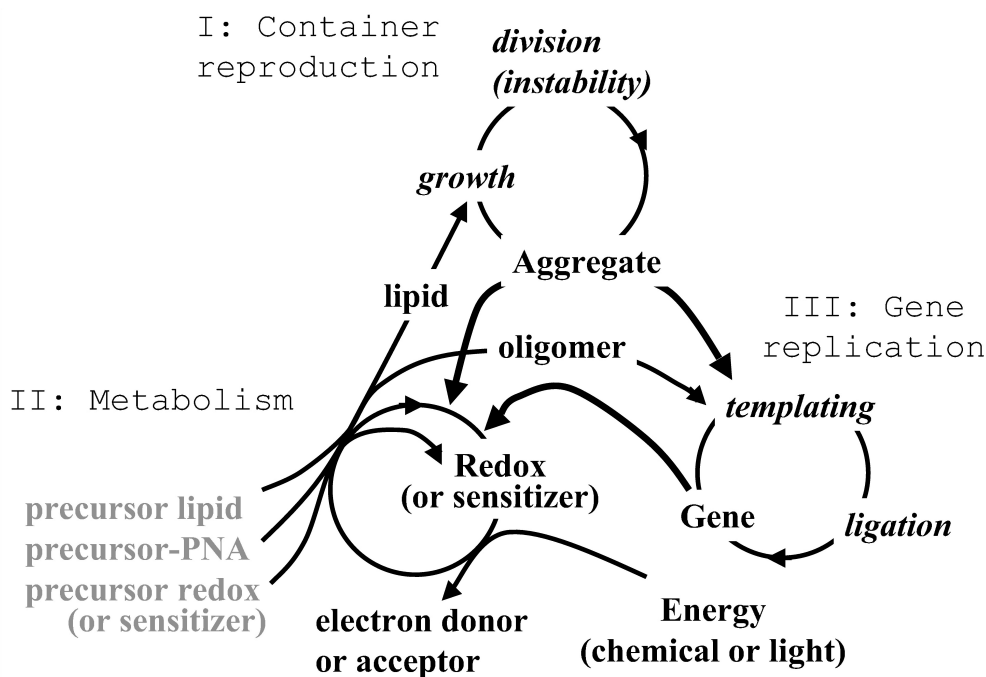


Figure 3. Causal structure of the full proto-organism discussed in Figure 2. Note the three aggregates with associated production cycles represented by light dotted circle arrows: container reproduction (I), metabolic cycle (II), and gene replication (III). The light solid arrows indicate production, and the heavy dashed arrows indicate catalysis. The (gray) precursor molecules constitute the food resources, and the whole system is fueled either by inorganic chemical or light energy.

block molecules), the efficient gene aggregate would eventually dominate, as a simple example of proto-organism selection.

Although this self-reproducing molecular aggregate does not constitute a modern cell, we believe it is a first model of a concrete molecular system that results from thermodynamically downhill processes and that combines container, metabolism, and proto-genes in a cooperative manner. Figure 3 summarizes the causal structure of the full proto-organism.

In closing the conceptual discussion, we review the proto-organism's key functional properties and how they are generated. This is a continuation of the discussion on dynamical hierarchies from [77], and it is not essential for the coming sections. It is clear that with a lipid aggregate we already have three natural levels of description with distinct observables: (1) the properties of the individual water molecules and monomers that make up the amphiphilic polymers, (2) the properties of the amphiphilic polymers (e.g., elasticity), and—for the amphiphilic aggregates—(3) an inside, which is hydrophobic, and an outside, which is mostly hydrophilic, the two parts having different chemical (catalytic) properties. Also, the aggregate has the ability to load a variety of different molecules, which we later return to in detail. As we add molecular objects to the lipid aggregate, the fundamental level of description (mainly the length

scale) does not change, but the observed properties of the composed structure may change dramatically. We see this happen when a proto-metabolic process (redox or photo-driven) is implemented, which enables the structure to grow and reproduce: As more lipid molecules are produced, the aggregate grows and eventually divides, which constitutes self-reproduction. We may ascertain the order of (emergent) functionality that is generated simply by paying attention to which of the substructures are necessary for the observed functionality. Then we may enumerate (order) each of the functionalities successively as they are generated (by the assembly of more subsystems), under the condition that each earlier functionality is a prerequisite of the next following functionality. In this context, for example, the proto-container (third order functionality) is needed before a redox-driven proto-metabolism (fourth order functionality, but still third level structure) can be established, and further, the proto-genes (fifth order functionality, but still third level structure) require the proto-metabolism to exist. However, our proposed photo-driven metabolism requires a particular electron relay chain to function properly, and this electron relay is implemented within the templating polymers. Thus, the proto-metabolism and the proto-genetic system may both be characterized as fourth (or fifth) order functionalities in this context.

3 Proposed Physicochemical Implementation of the Proto-organism

3.1 Assembly, Stability, and Self-reproduction of the Lipid Proto-container

It is important to note that the lipid aggregates in our approach are utilized in a somewhat different manner from most other model studies of the origins of life or artificial cell approaches. Our lipid aggregates have three key functionalities, and the first is unique to our approach.

1. As a proto-container, it holds together the other two key aggregates, the proto-metabolism and the proto-genes. However, it does so either at the exterior surface of the aggregate or deep within the lipid phase. The aggregate does not contain these other aggregates, that is, in the interior (water) volume of a vesicle, as it does in the typical lipid-world picture.
2. Also due to the container property, close spatial proximity results in very high local concentrations, which are key for many of the processes.
3. Finally, the interior lipid phase as well as the water-lipid interface have very different physicochemical properties from that of the bulk water phase, and so these three different physicochemical environments can be used to enhance certain reactions, in sequence. Both the lipid phase and the lipid-water interface act as catalysts.

Simple self-reproduction of micelles and vesicles from fatty acid molecules is demonstrated in the literature, in particular by Luisi and coworkers [47–50, 99]. It is also shown that long chain fatty acids are able to form bilayer structures in water, depending on the pH value and salt concentration of the solution; this phenomenon has been investigated by Deamer and coworkers [17–19, 1, 59]. For an example, vesicles are formed when the solution pH is equal to the pKa of the acid in the bilayer. On generating (introducing) more fatty acid molecules in a lipid aggregate, the micelles or vesicles will reach their maximum loading conditions (size), and eventually they will split into two smaller assemblies via budding. For our purpose it is important to correlate the structures of the fatty acids with their ability for micellation and vesicle formation and with their stability properties.

Based on extensive studies of self-assembly of chromophore-derived fatty acids by Whitten et al. [103], as well as work done by Deamer and coworkers [17–19, 1, 59], who have aggressively addressed the fundamental questions associated with simple fatty acid bilayer structures over the past years, we can summarize the following important facts about the self-assembling of simple fatty acids.

No vesicles are observed under any conditions using carboxylic acids with chain lengths shorter than 8 carbons, but all fatty acids from 8 to 12 carbons in length produced obvious vesicles. Shorter hydrocarbon chains form micellar structures. The minimum concentration of monocarboxylic acid necessary to form vesicles in aqueous solution is a function of chain length. The concentration for vesicle formation ranges from 130 mM for octanoic acid to 10 mM for dodecanoic acid. Carboxylic acids with chain lengths of 8 or more carbons form vesicles in the pH range near the pKa of their terminal carboxyl group. Otherwise they form micelles. It is found that the optimum pH for each chain length also varies slightly with the concentration of the acid in solution. Longer chain lengths required slightly higher pH ranges for stability. The addition of alcohols of the same chain length as the monocarboxylic acid forming the bilayer membrane dramatically increases the pH range of stability and decreases the concentration of fatty acid necessary for stable vesicles to form. For example, nonanoic acid requires 20 mM concentration of acid in the presence of 2 mM nonanol to form stable vesicles, which are stable at any pH above 6.5.

Vesicle membranes composed of nonanoic acid are rather impermeable to ionic solutes such as KCl, with a half-time of osmotic gradient decay of approximately 40 min. The vesicles are more permeable to polar solutes like glycerol, which has a half-time of 3 min. It is established that carboxylic acids having chain lengths of 8 or more carbons are able to self-assemble into stable vesicles within certain concentration and pH ranges. Depending on the chemical structures of the fatty acid and the chromophore molecule, the lipid phase of carboxylic acid vesicles are able to store hydrophobic substances in the aqueous solution up to millimolar range.

Two approaches are demonstrated in the literature to form fatty acid molecules inside lipid structures. The first approach involves a buffered water solution containing preformed vesicles, overlaid with a small amount of insoluble carboxylic anhydride [48]. Under those conditions, a significant increase of the hydrolysis rate of the anhydride can be observed with respect to a reference system that does not contain vesicles in the water phase. The second approach involves irradiating a dispersion of the photocleavable water-insoluble precursor didecyl-2-methoxy-5-nitrophenyl phosphate to form didecyl phosphate surfactant in the vesicles [50]. The phosphate-phenolate bond was selectively hydrolyzed and was released in the medium. Light microscopy was used to assess the presence of vesicles, which were generally in the range of 1–10 μm in size.

In summary: the experimental conditions for lipid self-assembly, stability, and self-reproduction are well known, although the specific loading capabilities of the aggregates for the proposed precursor molecules (see Sections 3.2 and 3.3) are not well known.

3.2 Lipid Proto-metabolic System (Photo-driven)

We now present a simple *lipid proto-metabolic implementation* as summarized in Figure 4. In this system a proto-gene—a PNA [65] assembly within an amphiphilic aggregate (proto-container)—encodes the metabolic production of the lipids as well as the precursor genes.

As we already discussed in Figure 2, initially amphiphilic monomers (e.g., carboxyl acids and alcohols) assemble into amphiphilic aggregates. A PNA strand (a proto-

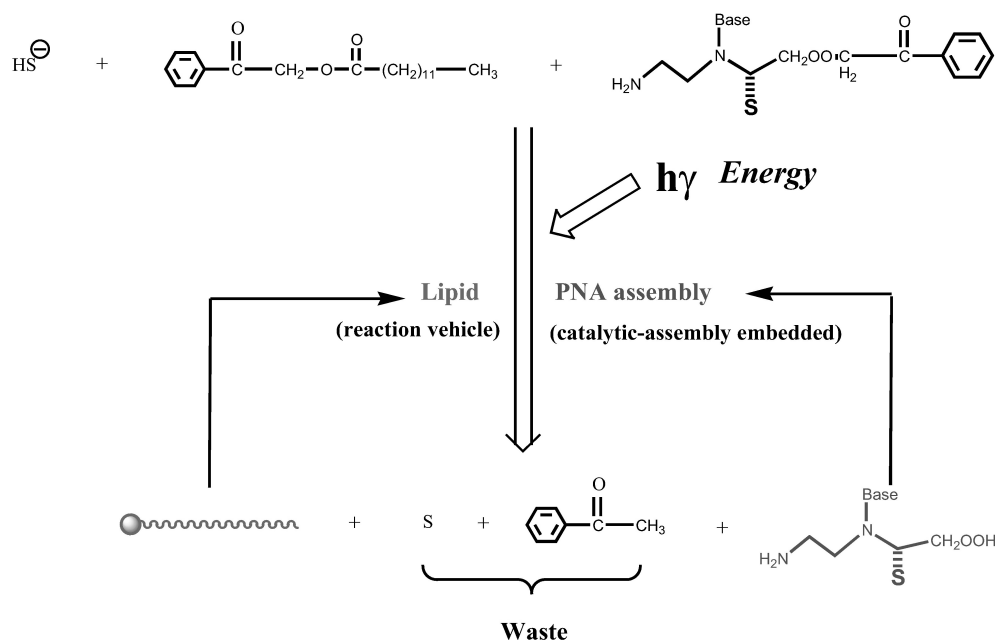


Figure 4. Simple physicochemical implementation of coupling between proto-genes, proto-container, and proto-metabolism. The metabolic chemistry in this system is based on two autocatalytic photosegmentation reactions where an ester is converted into a carboxylic acid derivative: one for the simple lipid and one for the proto-gene monomers (or oligomers). The PNA with attached sensitizer acts as a photocatalyst for both segmentation processes once the PNA is polymerized. Note that S symbolizes both free sulfur (end product after donation of an electron from sulfite) and the photosensitizer molecule attached to the PNA backbone.

gene), being hydrophobic, will attach to the lipid structure. Additional hydrophobic groups can be attached to the PNA backbone to ensure partition in the lipid structure as necessary.⁴ For simplicity we use PNA as a representative of such a class of templating polymers, without insisting on a particular backbone chemistry or nucleic acid alphabet. It is due to its favorable thermodynamic properties that PNA and not RNA is utilized in this system.

We propose a direct autocatalytic feedback between the PNA proto-genes and the production of both lipid molecules and more PNA precursor molecules. This feedback can be implemented by using a modified PNA as a photo-catalyst, in view of its charge transfer capabilities; see Figure 4. Charge transfer in DNA (with the same bases as in PNA) is well established experimentally, and theoretical models have been developed. We return to a detailed discussion of PNA charge transfer in Section 4.3.

In this proposed system, we use the photolysis of various phenacyl esters ($\text{PhCOCH}_2\text{-OCOR}$, where R denotes a long carbon chain) to produce surfactant molecules. It has been well established that phenacyl esters can undergo photo-induced C–O bond scission to form acetophenone (PhCOCH_3) and the corresponding carboxylic acid (RCO_2H , surfactant) in the presence of appropriate electron donor molecules [22]. However, due to the fast back electron transfer, the quantum efficiency is very low ($\sim 0.01\text{--}0.05$). A sensitizer coupled with an electron relay system can be introduced to block back electron transfer and thus increase the quantum yield of surfactant production; this technique has been successfully used in other systems. This provides an opportunity

⁴ Synthesis of PNA monomers with such backbones has been demonstrated by Peter Nielsen and coworkers and will be discussed in Section 3.3.

to achieve surfactant reproduction controlled by the proto-genes. We may incorporate a sensitizer molecule into the PNA backbone and use an appropriate sequence of nucleobases as the electron relay system. Accordingly, we will anchor the sensitizer (S) directly on the PNA backbone and use adenine (A) and guanine (G) as the effective electron relay unit.

We predict the following dynamics in this proposed system. Upon the excitation of S by irradiation, the excited state of S can be quenched by the phenacyl derivative to generate a contact ion pair of S cation radical and a phenacyl anion radical. Due to the close geometry of the PNA strand, the S cation radical can be efficiently scavenged by A to yield an A cation radical and the recovery of the sensitizer. The A cation radical can further react with adjacent G to yield a G cation radical, which is eventually quenched by an electron scavenger such as HS^- . Since the formed phenacyl anion radical and the G anion radical are far away from each other, back electron transfer will be avoided and the formed phenacyl anion radicals will have enough time to undergo the fragmentation reaction to produce the corresponding carboxylic surfactant molecules. Without the presence of the nucleobase sequence as the electron relay unit, the back electron transfer between the sensitizer cation radical and phenacyl anion radical will dominate the reaction, and the yield of production of surfactant will be low. Therefore, surfactants can be replicated only when a certain base sequence is present. The reaction can be easily followed by nuclear magnetic resonance (NMR) and Fourier transform infrared (FTIR) observation of the formation of carboxylic surfactant, as well as by a transient spectroscopic study monitoring the formation of A and G cation radicals.

We may apply the same chemistry to demonstrate the production of PNA repeat units (Figure 4) by converting amino ester to amino acid monomer. Under these conditions, precursor PNA dimer (or, alternatively, oligomer) will be effectively converted to functional PNA dimer (or oligomer), which subsequently can be polymerized (the use of PNA dimer instead of monomer will diminish the cyclization reaction of PNA monomer).

To summarize the proposed photo-metabolic system: in the presence of adenine, guanine, and sensitizer modified PNA (an electron relay system), a photoinduced electron transfer reaction occurs between the sensitizer and the phenacyl ester; consequently, the sensitizer cation radical can receive an electron from adenine, forming an adenine cation radical. This adenine cation is then reacted with adjacent guanine to yield a guanine cation radical, which is eventually scavenged by HS^- . The amphiphilic molecules can only be synthesized when the particular AG base sequence is present. As a result of the increased production of amphiphilic (lipid) molecules, the aggregates grow, become unstable, and eventually divide as discussed in Section 3.1. By using the same photofragmentation process, precursor PNAs (dimers or oligomers) form functional PNA oligomers, which can hybridize with a complementary PNA template, and this double-stranded PNA complex determines the photocatalytic properties. This autocatalytic feedback system establishes our first goal of coupling the proto-container, the proto-metabolism, and the proto-genes. However, note that this system does not yet constitute a full proto-organism as defined in the last section. There is no template directed replication of the proto-genes, which will be introduced in Section 3.3.

Our preliminary experiments support the above approach. As an experimental model system, we use an amino-pinacol derivative as the electron donor molecule and a phenacyl ester as the electron acceptor molecule. As depicted in Figure 5a, the excited state of the pinacol can be quenched by the phenacyl derivative (with quenching constant of 10^2 M^{-1}) to generate a contact ion pair consisting of a pinacol cation radical and a phenacyl anion radical. However, due to the fast back electron transfer reaction rate, 98% of the contact ion pairs return to the starting materials. A small fraction of contact ion pairs undergo charge separation, which eventually leads

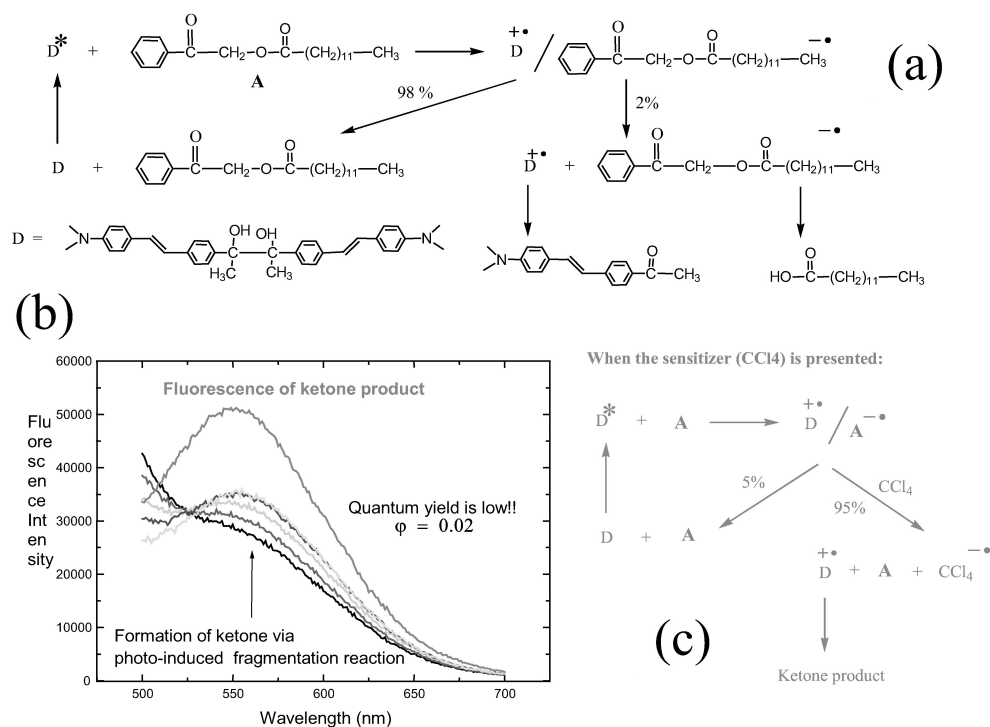


Figure 5. Photoinduced fragmentation reaction of an amino-pinacol derivative as the electron donor molecule and a phenacyl ester as the electron acceptor. (a) The excited state of the pinacol can be quenched by the phenacyl derivative to generate a contact ion pair of a pinacol cation radical and a phenacyl anion radical. (b) From the fluorescence spectra we can deduce that the product quantum yield is only about 2%, due to the dominant back electron transfer processes [recall (a)]. (c) CCl_4 can efficiently intercept the phenacyl ester anion radical inside the contact ion pair to generate a CCl_4 anion radical. Thus about 95% of contact ion pairs will now undergo charge separation and lead to a high quantum yield of product formation. See text for details.

to the formation of ketone molecules from the pinacol cation radical and of carboxylic acid from the phenacyl ester anion radical. The formation of ketone molecules can be easily monitored by their fluorescence (Figure 5b). From the fluorescence spectra, we can deduce that the product quantum yield is only about 2%, due to the dominant back electron transfer processes. The quantum yield of the ketone product can be increased significantly when the same reaction is carried out in the presence of a high concentration (5 M) of a second electron acceptor, such as carbon tetrachloride (CCl_4). As shown in Figure 5c, CCl_4 can efficiently intercept the phenacyl ester anion radical inside the contact ion pair to generate CCl_4 anion radical. Thus 95% of contact ion pairs underwent charge separation and led to the high quantum yield of product formation, as demonstrated in the fluorescence spectra. This proves our point that an electron relay system can dramatically enhance the photofragmentation yield, although we would need to introduce an electron donor and not an electron acceptor as is used here. Also, since our example is a homogenous reaction, a high concentration of CCl_4 is essential to react with the contact ion pairs; while in the heterogeneous case, with proper molecular alignment such as attachment to the PNA backbone, a low concentration of sensitizer should achieve the same effects. However, this has not yet been proven experimentally for this particular system.

3.3 Lipid Proto-gene Subsystem

To implement a template directed ligation of PNA in a lipid aggregate with a subsequent replication process, the above systems (Sections 3.1 and 3.2) have to be expanded. Initially, note that it is thermodynamically highly unfavorable to attempt a ligation in water. Assume, as in Figure 2, that we have lipid aggregates with PNA strands (templates) attached at the surface. If this system is supplied with two pieces of PNA complementary to the template, a hybridization reaction between the original template and two complementary oligomers will occur, as a consequence of thermodynamics. As this new, three-component complex sits in the hydrophobic environment, a ligation (polymerization) is possible, driven by the expulsion of water from the lipid layer (by le Chatelier's principle). The possibility of polycondensation of amino acids in the lipid phase has already been experimentally demonstrated by Luisi and coworkers [5]. The ligation reaction may be further enhanced either by cyclic drying of the lipid aggregates or by catalysis. Simple catalysts may also speed up the kinetics, as in a general base catalysis using pyridine or a Lewis acid catalysis using a transition metal.⁵

The next step in the template-directed replication of PNA in the lipid aggregate requires a reasonable equilibrium to be established between the hybridized (double stranded) and single stranded PNA molecules. This is necessary to enable further templating reactions. Experimentally this can be established by variations in temperature or pH, but employment of modified PNA backbones that decreases hybridization energies (such as the β -alanine backbone [37]) is also an option.

The result of this proposed process is a multiplication (replication) of the proto-genes. However, this lipid-catalyzed PNA replication process has not yet been experimentally demonstrated.

Preliminary experimental results that support the above approach are available from Nielsen and coworkers. PNA-PNA and PNA-DNA melting curves in water have already been characterized extensively [79, 90]. PNA-directed chemical synthesis of another sequence-complementary PNA strand has been demonstrated in the aqueous phase using a C₁₀ template onto which G dimers were oligomerized via carbodiimide activation [11]. Furthermore, PNA monomers of thymine in which the backbone glycine is replaced with phenylalanine, isoleucine, or valine are described in [33, 73], and it has been demonstrated that these monomers can be incorporated into PNA oligomers without major loss of hybridization potency. Based upon the structure of aminoethylglycine PNA, we expect that it will only partition slightly into the organic phase, which indicates that the addition of a hydrophobic group to the backbone might be necessary. How much extra backbone hydrophobicity is necessary is an experimental question that could be investigated with detailed MD and MD lattice gas simulations.

Template directed ligation has been experimentally demonstrated in water with energized dimers [11], but it has not yet been shown in a lipid phase. However, if polymerization of PNA is achieved in lipid layers, one can turn to the more difficult problem of replication of PNA molecules. This process requires that a strand of PNA serve as a template for the assembly of a second complementary strand of PNA. A complex between a template strand of PNA and two oligomers might form in either the aqueous phase or the lipid phase. The condensation reaction would occur within the lipid phase, driven by the expulsion of water from the hydrophobic environment within the lipid layer. True replication requires that, once a double-stranded PNA decamer is formed by template-directed synthesis, the decamer can dissociate and both strands then serve as templates for another round of replication. The best choice of PNA (with respect to backbone and base composition) will eventually have to be determined from theoretical calculations and the experiments described above, which will help choose

⁵ Originally proposed by Shelley Copley.

molecules that form metastable complexes that can separate to provide templates for subsequent rounds. Identification of appropriate conditions for metastable PNA complexes is probably the most uncertain experimental task.

Initially, in order to facilitate the ligation, it is possible to study the reaction using PNA pentamers with an activated (through an active ester function, NHS-pentafluorophenyl) carboxyl group. By this approach one can bypass the uphill thermodynamics of the ligation and study the issue of multiple turnover in more detail. Again, analogous experiments in aqueous solution have already been performed and published [11].

These proposed experiments will provide insight into a problem that has vexed prebiotic chemists for decades. It long has been recognized that the condensation reactions required to form polymers such as nucleic acids and proteins are thermodynamically unfavorable without strong activating groups, but likely activating groups and conditions for driving the endothermic formation of activated monomers have not been identified. The special environment provided by a lipid layer could certainly promote these types of reactions.

3.4 Proto-lipid, Proto-metabolic, and Proto-genetic Integration

In order for PNA to qualify as a true proto-gene, it has to include *genomic* information that *encodes* the production of parts of the proto-organism. Further, the genomic information has to be inheritable so that the newly generated daughter proto-organism has a similar genome to the parent. Here, we propose a proto-organism where the replication is controlled by a catalytic unit having a particular base sequence embedded in the PNA, and we go through the predicted dynamics of the full assembly. Since only certain sequences of nucleobase pairs can catalyze the production of the amphiphilic molecules, the replication rate of the lipid aggregate will be determined by the PNA. Given our ability to control the surfactant reproduction through certain nucleobase pairs, we should be capable of using PNA as a genomic code to guide the replication of surfactant assemblies. As proposed in Figure 6, we may start with a surfactant self-assembly (a supported bilayer, a micelle, or a vesicle) containing a PNA oligomer (hexamer) with a TCTCTC base sequence and sensitizers S attached to the C monomers. The synthesis of PNA monomers with a sensitizer attached to the backbone can be achieved by following the reaction scheme developed by Nielsen and coworkers. (A simpler implementation scheme, which does not involve the integrated PNA and sensitizer, is discussed in Figure 7.)

Now all the components for the production of surfactant and PNA dimer as well as the PNA polymerization catalyst are present in the lipid phase. However, since no A-S-G complex is yet available, the surfactant replication process is inhibited by the back electron transfer from the sensitizer cation radical and phenacyl anion radical, as mentioned earlier. The replication process will eventually start with the formation of functional PNA dimers (or oligomers), which are able to polymerize, as outlined in the previous section. While the formed PNA dimers (or oligomers) are floating around in the lipid phase, the longer PNA template will attract the complementary PNA dimers (AG, AG, AG) to form a duplex. Consequently, these dimers will align themselves according to the sequence of bases on the PNA template. Since efficient A-S-G complexes are available now, they will speed up the co-sensitization process for the photofragmentation reaction. Consequently, surfactant replication will be initialized. The production of surfactant will increase as the alignment of A, S, and G along the template PNA is increased, since the sensitizers (S, A, and G) are not consumed in this case. Meanwhile, template-directed ligation and synthesis of PNA (perhaps in the presence of the catalyst) will also start, and a complementary PNA oligomer AGAGAG will be formed. Under certain conditions, the double stranded PNA is subject to dissociation

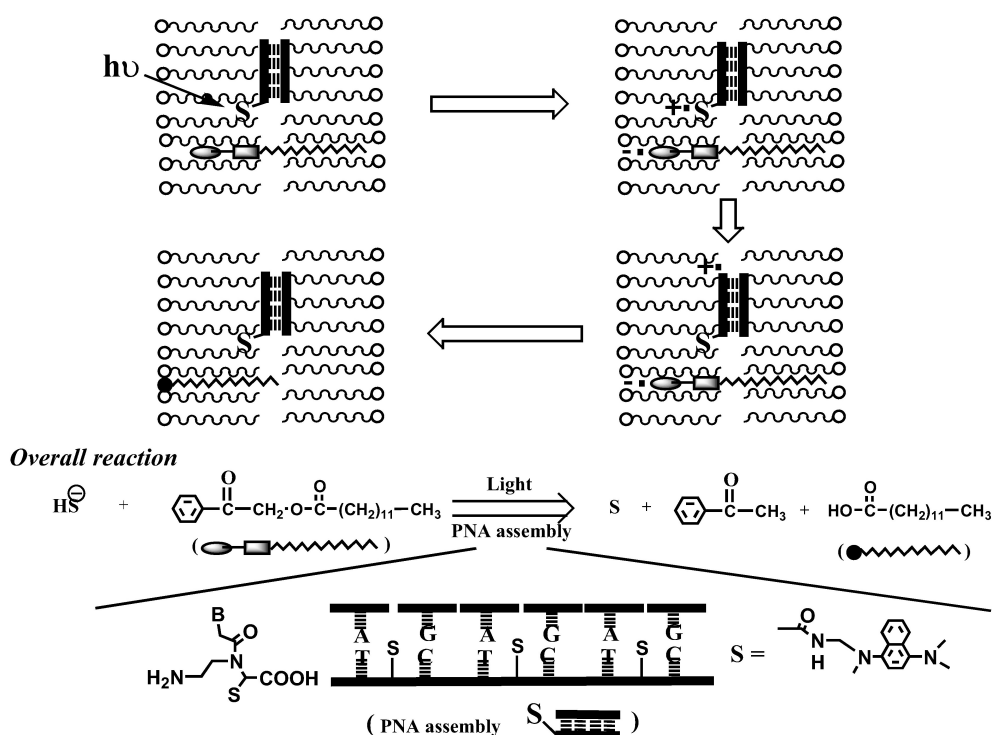


Figure 6. Scheme for proto-organism replication. The top left tableau depicts the excitation of the sensitizer due to an energy rich photon. The energized sensitizer causes charge separation between the sensitizer and the lipid precursor (penacyl), in the upper right tableau. The middle right tableau depicts the neutralization of the sensitizer by the delivery of an electron from the PNA electron relay chain. The precursor lipid molecule is still energized and will eventually break (fragment) into a lipid (carboxyl acid) and a phenyl group, which is depicted in the middle left tableau. The breakage occurs at the ester bond. The overall reaction is summarized in the lower part of the figure, where the structure of the sensitizer-enriched PNA is also indicated. Note that only double stranded PNA has photocatalytic properties. Also note that S means sulfur in the overall reaction, but means the sensitizer molecule attached to the PNA backbone in the other tableaux. See text for details.

into two single strands,⁶ while at the same time the continued production of surfactants from the hydrophobic precursors embedded in the aggregate eventually causes the aggregate to become unstable and split into two aggregates. The newly formed surfactant assemblies embedded with PNA templates are subject to further replication to form even more assemblies as more PNA and lipid precursors are supplied. At the final stage, the population of the proto-organism assemblies containing the favorable sequence of PNA template will predominate.

By integrating this template-directed replication process with the above template-encoded lipid and template production processes, a cooperative (autocatalytic) feedback has now been established between the three key elements in the proto-organism: (a) proto-container, (b) proto-metabolism, and (c) proto-genes, as we originally set out to do.

This autocatalytic feedback also forms the basis for Darwinian evolution of the system. For example, starting with two different proto-genes on different lipid aggregates, one template with and one without AG components, will result in an evolutionary takeover by the aggregates with AG-rich strings, because it facilitates the production of (proto-container) lipids. Due to the predicted parabolic replication kinetics (Sec-

⁶ The PNA disassociation process (balance) is extensively discussed in Section 4.4.

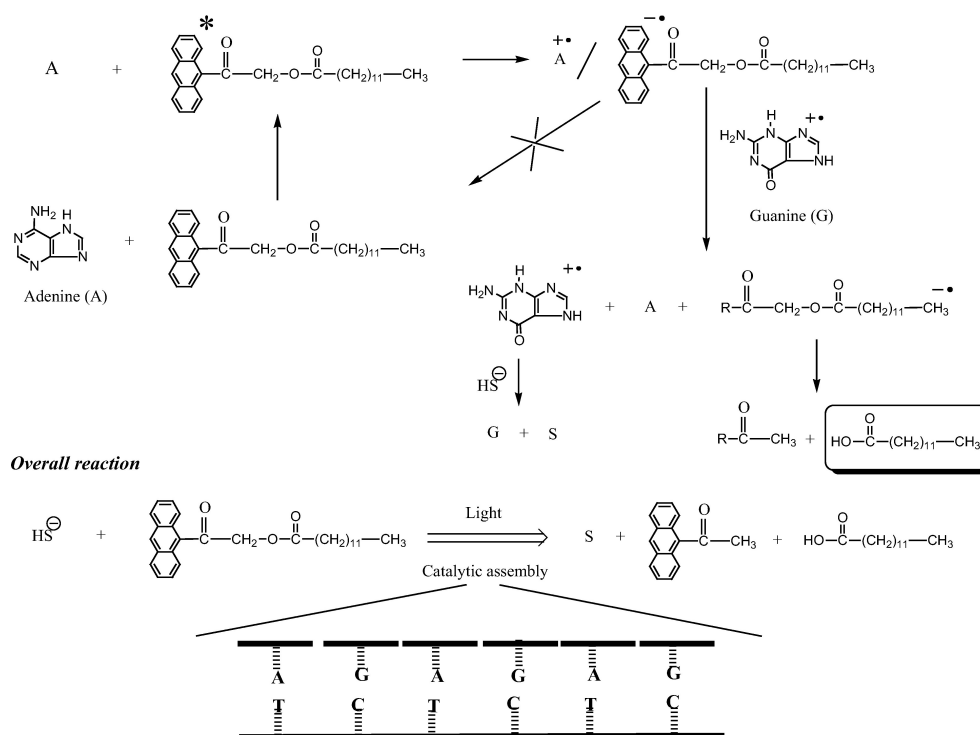


Figure 7. Alternative physicochemical implementation of the photo-driven segmentation reactions without direct attachment of the sensitizer to the PNA backbone. This approach could be used as an alternative to the PNA monomers with the sensitizer attached. The overall scheme is essentially the same as discussed in Figure 8, although sensitizer molecules have to be provided as part of the resources.

tion 4.4), these aggregates could also carry (parasitic) coexisting genes that can replicate, but cannot function as efficiently as an electron relay.

Although this proposed self-reproducing molecular aggregate, or proto-organism, does not constitute an actual cell, to our knowledge it is the first model of a concrete molecular system that results from thermodynamically downhill processes and that combines container, metabolism, and genes into a cooperative structure. However, to obtain a fully integrated and operational experimental system is a very difficult task, even if each of the subsystems works in isolation. We cannot exclude that as yet unknown emergent properties make it impossible. We don't expect this will occur, but it is a question that eventually has to be addressed experimentally. The full causal structure of this proto-organism is discussed in Figure 8. For completeness we discuss an alternative physicochemical implementation of the proto-organism processes based on a redox metabolism, although in less detail.

3.5 Alternative Implementations (Redox-Driven)

In Figure 9 the production of both surfactant and functional PNA molecules is driven by chemical energy sources. Head groups, such as carboxylic or sulfonic acids, that are commonly present in amphiphilic molecules are usually synthesized by using harsh oxidation conditions such as light. Alternatively, amine groups, which can also act as hydrophilic head groups, are often synthesized through a mild reduction process. For example, organic nitro compounds can easily be reduced to amines. Many reducing agents (the most common ones being zinc, tin, and iron cations), as well as catalyzed

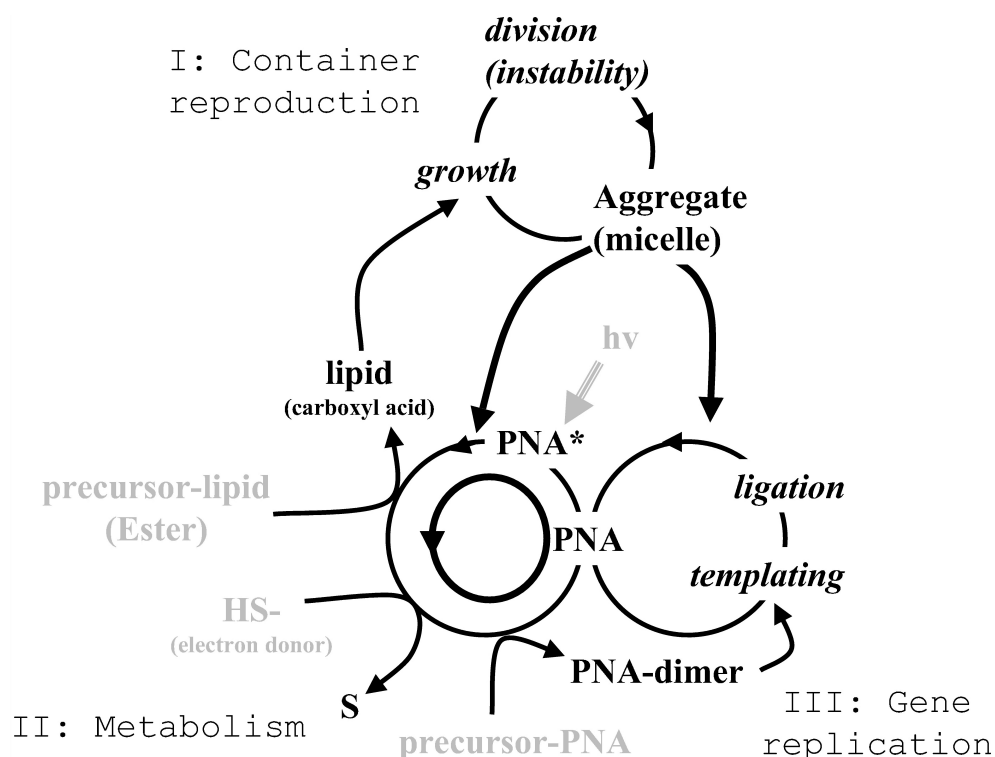


Figure 8. Causal structure of the simple physicochemical implementation of the full proto-organism discussed in Figure 2 and in the text. Compare with Figure 3. Again the resources for the structure are indicated in gray: precursor lipid (a hydrophobic ester), precursor PNA, and an electron donor (sulfite), as well as photoenergy (light gray). Also note the three main physicochemical cycles: the aggregate growth and division cycle (I) fed by the production of lipids; the metabolic (PNA) cycle (II) fed by light (note how the PNA exists in a neutral form and an energized form PNA*); and the PNA replication cycle (III) fed by short (pre-template) PNA strands. Both these cycles are kept together and catalyzed by the lipid aggregate (heavy dashed arrows), but in this system the proto-metabolic and the proto-genetic molecules are identical. Thus, the gene replication cycle (III) and the metabolic cycle (II) are linked more tightly than in Figure 3.

hydrogenation processes, can reduce nitro compounds. One such nitro reduction reaction is the so-called Zinin reduction [72]. This reaction uses sulfides or polysulfides, which actually are plausible prebiotic reducing agents, since the presence of HS^- or H_2S is extensive in volcanic geothermal environments. This process may occur through the formation of single chain lysophosphatidyl ethanolamine (lysoPE), a derivative of a natural phospholipid molecule. It has been reported that single chain phosphates assemble into closed bilayer vesicles, which can be easily observed by optical or electron microscopy [34]. We may use phosphatidyl nitroethanol as the precursor reactant and utilize chemical energy from HS^- to drive the reduction process, as shown in Figure 9. Since the HS^- is located in the aqueous solution and the nitro phosphates will reside in the membrane (organic) phase, we may incorporate appropriate electron transfer compounds such as dinitrophenolindophenol or phenazine methosulfate in the membrane to facilitate the redox reaction. Standard chromatographic methods may be used to monitor the formation of the single chain phosphatidyl ethanolamine, and products can be verified by electrospray mass spectrometry, NMR, and FTIR. We also expect that as the synthesis occurs, the newly formed lysoPE will incorporate into the existing bilayer vesicles to produce growth.

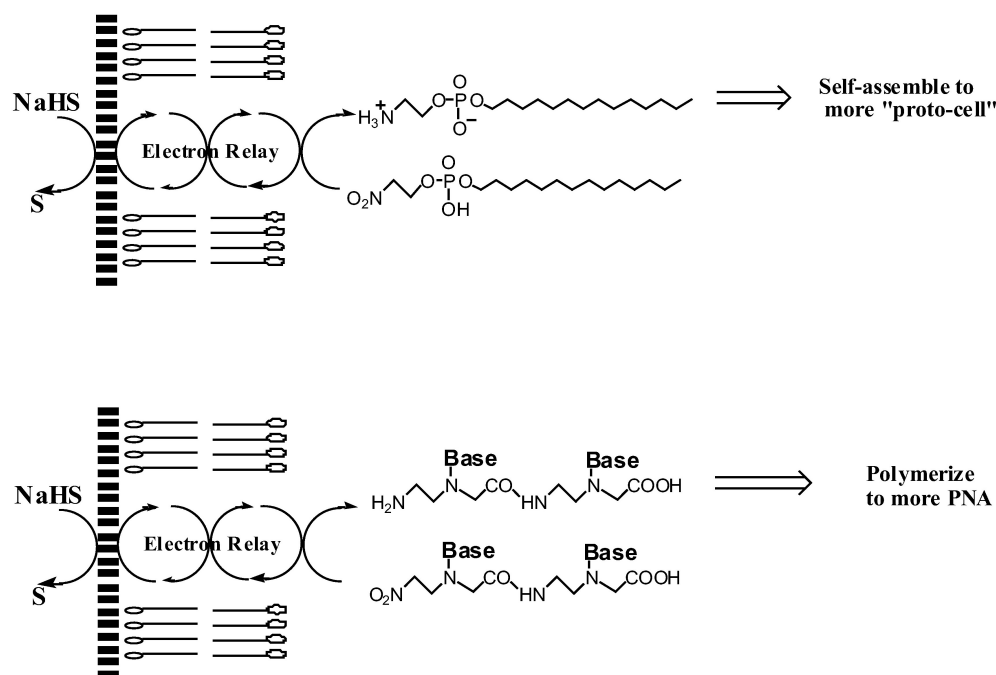


Figure 9. Reductive synthesis of surfactant and PNA dimer in a vesicle system. See text for details.

In order to demonstrate the production of functional PNA repeat units, a nitro derivative of PNA dimer may be used as the precursor for the synthesis of PNA dimer.⁷ Due to the hydrophobicity of the PNA precursor, it will reside in the vesicle membrane (organic) phase as originally discussed in Figure 2, while the HS^- ion is located in the aqueous solution. Therefore, an electron relay system may be employed to ensure the electron flow between the energy source and the precursor. Under these conditions, the nitro group will be effectively reduced to the amine group and thus the functional PNA dimer will be formed.

By blending all the above components in the lipid phase, the proposed system is capable of self-generating more surfactants and PNA dimers, both of which are able to assemble to more organized structures (vesicles) or to polymerize to more PNA oligomer. Since the total free energy decreases as the surfactants assemble and the dimers at the same time polymerize, this slight energetic drive will lead to a coupling of the two replication processes, although this coupling is much weaker than the metabolic coupling discussed in the previous sections. Spatial proximity will also act as a stabilizing factor.

So far this scheme does not account for the production of more redox molecules (such as dinitrophenolindophenol or phenazine methosulfate), so the proto-organism has to be on *life support* in the sense that these redox molecules have to be part of the available resources. The same scheme can be used for the proto-gene replication, but it is not clear how the encoded base sequence can directly influence (catalyze) the redox kinetics. It is, however, conceivable that the kinetics will depend on the specific properties of the PNA strand (e.g., chemical or folding), but without direct experiments we do not dare to make any specific predictions.

⁷ Again, the reason to use PNA dimer instead of monomer is to diminish the cyclization reaction for the monomer.

In the previous sections we have been focusing on a light-driven proto-metabolism, due to its simplicity in the laboratory context, but light may not have been the energy form utilized at the origins of life. Evidence elsewhere seems to indicate that the surface of the early Earth was too harsh an environment due to the late accretion bombardments. The protected pore space in the subsurface (in the planetary crust) may very well have provided the richest and most appropriate chemistry with water, hydrocarbons, complex organics, minerals, and interfaces on the young Earth. In addition, phylogenetic studies based on RNA sequences seem to indicate that the redox-driven metabolism is earlier than the photo-driven metabolism. This all points to a subsurface redox-driven origin of life [15].

4 Thermodynamic and Kinetic Aspects of the Proto-organism

The key thermodynamic and kinetic issues associated with the proto-organism can be divided into five main problems: (a) the assembly of the lipid aggregates, (b) the association (loading) of the proto-genes and the metabolic precursor molecules in the lipid aggregates, (c) the thermodynamics and kinetics of the metabolic processes, (d) the thermodynamics and kinetics of the template directed ligation and replication processes, and (e) the thermodynamics and kinetics of the full system replication.

Under standard conditions (constant pressure and temperature) the change in Gibbs free energy, ΔG , for a given physicochemical process is defined as

$$\Delta G = \Delta U + P \Delta V + \sum_i \mu_i \Delta n_i - T \Delta S = \Delta H - T \Delta S \quad (1)$$

where ΔU essentially amounts to the overall change of internal (molecule to molecule) potential energy, ΔS is the internal entropy change, P is the pressure and T is the absolute temperature, and we assume that we can ignore volume changes ΔV . Since the system involves chemical reactions through the metabolic processes (breakage and formation of covalent bonds), additional terms $\sum_i \mu_i \Delta n_i$ are needed to take account of the changes in ΔG due to the changes in the concentrations of chemical species. Δn_i denotes the concentration change in species i , and μ_i the chemical potential corresponding to species i . The enthalpy ΔH is just shorthand for the first three terms. The thermodynamic potential change ΔG is discussed in connection with the key proto-organism processes in Figures 10, 14, and 15. Although most of the proposed subprocesses involved in the proto-organism dynamics are well known from earlier experiments in different contexts, very little is known about the integration of these reactions, which is necessary to form the full life cycle of the proto-organism.

Recall that thermodynamics does not involve any time scales. The thermodynamic properties are determined by the equilibrium conditions for the system, including the equilibrium concentrations for the physicochemical species involved. The system time scales, as determined by the kinetics, are generated by the details of the assembly and reaction mechanisms. The time scales are influenced by catalysis, but the thermodynamic conditions are not. There is a clear connection between the thermodynamic conditions and the kinetic reaction constants in that reaction constants are defined in terms of ratios between the thermodynamic equilibrium concentrations.

4.1 Thermodynamics and Kinetics of the Lipid Aggregation Processes

Self-assembly of lipids in water is an extensively studied area, which includes experimental and theoretical work by our team. Self-assembly fundamentally involves

microscopic processes of the form



where $A(n)$ is a lipid aggregate of size n (consists of n lipid molecules). One special case of Equation 2 is important: $n = 1$, $m \gg 1$, as it characterizes the assembly and disassembly of individual lipid molecules into a micelle or vesicle. Equation 2 also accounts for the fusion of smaller aggregates and the breakdown (division) of larger aggregates. Breakdown of larger components into multiple components (more than two) can be described through iteration of Equation 2.

In this subsection we examine three interrelated theoretical approaches to the molecular self-assembly of aggregates: (a) thermodynamics, (b) kinetics, and (c) detailed 3D simulations, in increasing order of the typical mathematical complexity.

Lipid self-assembly has two thermodynamic drivers. One driver, perhaps the easier to understand directly, is a decrease in ΔU , the potential bond energy, through self-assembly. Through self-assembly the hydrophobic tails of the amphiphilic molecules are pushed together and out of the way of the water molecules so that more water-water bonds can form. This yields lower internal potential energy, because the water-water hydrogen bonds are stronger (≈ 20 kJ/mole) than the water-hydrophobic dipole-induced dipole bonds (≈ 2 kJ/mole). Thus, lipid self-assembly has a tendency to minimize the interface between the water molecules and the hydrophobic molecules.

In addition, lipid assembly has an entropy driver, which is actually stronger in absolute terms than the energy driver. This driver is due to an overall increase in the degrees of freedom for all molecules as the total water-lipid interface area decreases. The water-lipid interface generates a layer, or network, of relatively stable water-water bonds. This happens because the water next to the hydrophobic lipid interface has fewer choices to build strong water-water hydrogen bonds. Thus, an entropy gain ΔS is generated in lipid self-assembly. The entropic part of the process is often referred to as the *hydrophobic effect*. Together the entropy gain and the decrease in internal energy yield a negative (Gibbs) free energy change, $\Delta G < 0$; recall Equation 1.

In the following we discuss a variety of approaches to uncover the nature of lipid self-assembly processes and the specific properties of the generated aggregates, such as the size distribution, stability, and the aggregate dependence on the physicochemical parameters. A deeper understanding of the proto-container dynamics (as well as the dynamics of the other subsystems) is necessary prior to the assembly of the more complex, full proto-organism.

The free energy change in the micellation process can be determined from the equilibrium between the concentration A of n free monomers and the concentration A_n of the n -aggregate (micelle of size n) (recall Equation 2), using the fundamental relation

$$\Delta G = RT \ln K, \quad (3)$$

where R is the gas constant and the equilibrium constant is defined as

$$K = A_n / A^n. \quad (4)$$

Substituting Equation 1 into Equation 3 and rearranging the terms, we get the well-known relation

$$\ln K = -\frac{\Delta H}{R} \frac{1}{T} + \frac{\Delta S}{R}. \quad (5)$$

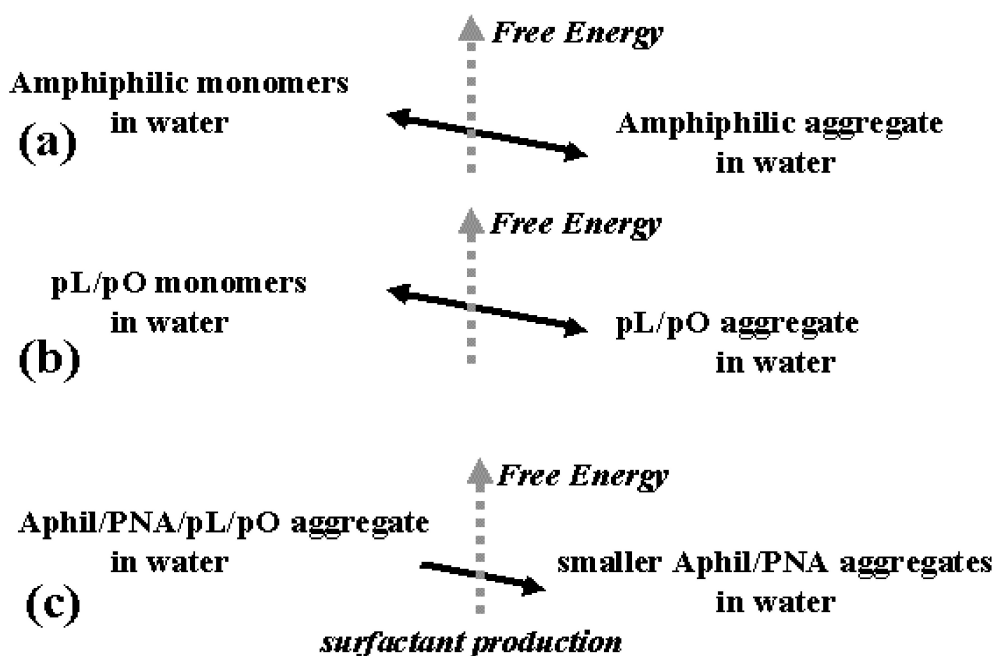


Figure 10. The thermodynamics of (a) the lipid self-assembly, (b) assembly (phase separation) of the hydrophobic precursor lipids (and a few PNA precursors) in water, and (c) the association (loading) of many precursor lipids and one or a few proto-genes and precursor oligomers into the lipid aggregates, following lipid production. The vertical axis indicates the relative free energy level of each component. (a) and (b) both are reversible reactions, and so is the loading of the precursor molecules into the aggregates, whereas the metabolic production of lipids and functional PNA from the loaded precursor molecules (c) is irreversible. Note that the relative free energy levels depend on both the detailed chemical species and their relative concentrations on each side of the reaction arrow. See text for details.

Thus, a plot of $\ln K$ as a function of $1/T$ yields ΔH and ΔS from the line's slope and intercept, respectively. Experimentally this relation can be determined by measuring K at different temperatures.

Lipid phase diagrams are usually very complicated. Figure 10a discusses general lipid self-assembly, and micellation is perhaps the simplest relevant example of such a process in the context of the proposed proto-organism. As a rule of thumb, the micellar structure may be derived from three simple parameters defining the lipid molecules: (a) the hydrocarbon chain length l_c , (b) the head group size Δ , and (c) the head group area a_0 , defined as part of the surface of a micellar aggregate [81]. Obviously, we can define the total micellar surface area A and the (internal) hydrophobic micellar volume V_c as

$$A = n a_0 = 4\pi(l_c + \Delta)^2 \quad \text{and} \quad V_c = \frac{4}{3}\pi l_c^3 \quad (6)$$

respectively, where n is the aggregation number (micellar size). Empirical studies show [81] that we may approximately predict the resulting aggregates as follows:

$$\frac{V_c}{l_c a_0} \begin{cases} = 0-0.33 & \text{generates spherical micelles in water} \\ = 0.33-0.5 & \text{generates rods in water} \\ = 0.5-1.0 & \text{generates lamellar structures in water} \\ > 1.0 & \text{generates reverse micelles in nonpolar media} \end{cases} \quad (7)$$

but the kind of aggregate also depends on the lipid concentration. These relations may be useful in that we may assume that the simple proto-container changes between spherical micelles and rod-shaped structures as we move between pure lipid aggregates and loaded aggregates containing larger amounts of precursor lipids; see Figure 10c. However, Equation 7 is only a simplified version of the story.

Lipid self-assembly has a rich dynamics, and the complex higher order structures of the assembled lipids depend on the properties of the specific lipid molecules. The assembly structure depends on the number and length of the hydrocarbon chain(s) (m molecules of the type $-\text{CH}_2-$), the charge and structure of head group (W), and the properties of the solvent, which is normally water, together with, for example, the pH, salt concentration, and ion mixture. A typical lipid molecule is about $\frac{(m+2)}{3}$ Å in length ($\approx 2\text{--}4$ nm), where m is the number of midtail hydrophobic monomers. Up to a certain critical lipid concentration (the critical micellation concentration, or CMC), no micelles can form, but at concentrations above the CMC, micelles will start to assemble. The CMC strongly depends on the lipid and the solvent. Many micelles start forming in water at concentrations between 10^{-2} and 10^{-4} M, with the higher CMC for nonionic amphiphiles such as carboxyl acids [81]. The micellation time—the time it takes an initially randomly distributed solution of lipid molecules to assemble into micelles—is about 10^{-6} s.

The contribution from each lipid molecule to the change of free energy on assembling into a micelle may also be factored out into the different components (monomers) of the lipid molecule:

$$\Delta G_{\text{mic}} = \Delta G_{\text{mic}}(W-) + m \Delta G_{\text{mic}}(-\text{CH}_2-) + \Delta G_{\text{mic}}(-\text{CH}_3) \quad (8)$$

where $\Delta G_{\text{mic}}(W-)$ indicates the change of free energy due to the hydrophobic head group of the lipid, $m\Delta G_{\text{mic}}(-\text{CH}_2-)$ the change due to the m hydrophobic midtail hydrocarbon groups, and $\Delta G_{\text{mic}}(-\text{CH}_3)$ the change due to the hydrophobic end-of-tail hydrocarbon group [81]. The free energy gain associated with a typical (simple) lipid is about -30 kJ/mol (-300 meV/lipid molecule), which means that each $-\text{CH}_2-$ monomer contributes with about -2 kJ/mol (~ -20 meV/hydrocarbon monomer). The head group's contribution is small. For comparison the thermal energies at 300 K ($\approx 25^\circ\text{C}$) are about 25 meV. Typical aggregate sizes n for non-ionic micelles range from about 25 to several hundred lipid molecules.

To experimentally determine the micellation size n , we can define M as the concentration of monomers in n -aggregates at equilibrium, and we can define M_0 as the total concentration of monomers. Then $M_0 - M$ expresses the concentration of free monomers, and Equation 4 can be rewritten as

$$\ln M = n \ln(M_0 - M) + \ln K \quad (9)$$

It turns out that both M_0 and M can be determined directly from experimental observables [58] as long as the micellation causes a change in the absorbance spectrum, which it usually does:

$$M_0 = E_{\text{mice}}KL \quad \text{and} \quad M = \frac{R_a(I_1 - R_m I_2)}{R_a - R_m} \quad (10)$$

where E_{mice} is the extinction coefficient at the absorption maximum for pure micelle; K is again the equilibrium constant; L is the length of the optical path; R_a is the ratio of micelle absorbance at λ_1 to that at λ_2 , which can be obtained from the pure micelle spectrum; R_m is the ratio of the monomer absorbance at the same two wavelengths λ_1

and λ_2 , which can be obtained from the pure monomer spectrum; and I_1 and I_2 are the individual absorbencies of micelles at λ_1 and λ_2 at any given concentration M_0 . The slope of the plot of $\ln M$ as a function of $\ln(M_0 - M)$ will give the average micellation size based on the Equation 9.

The details of the aggregation size distribution are not fully understood and depend strongly on the properties of the individual lipid molecules and the physicochemical conditions, that is, on how the individual lipid molecules join and separate in solution, how lipid molecules join and leave existing micelles, how micelles becomes unstable and split, and how smaller micelles join to form larger micelles (recall Equation 2). As a simple example we may assume the detailed lipid aggregate kinetics to consist of only three processes, ignoring the assembly of larger aggregates:



describing how lipids join and leave micelles, how larger micelles divide into two smaller micelles, and how micelles break apart. Here the parameters k , s , and d are assumed known functions of the aggregate size. Under the assumption that we can express the size of a lipid aggregate as a continuous variable x , it can then be shown [95] that the dynamics of the probability $P(x, t)$ of finding aggregates of size x at time t is determined by

$$\begin{aligned} \partial_t P(x, t) = & -\partial_x[k(x)P(x, t)] - s(x)P(x, t) \\ & - e(t)d(x)P(x, t) + \int_x^\infty T_x(y)s(y)P(y, t) dy \end{aligned} \quad (12)$$

where ∂_t and ∂_x are the time and aggregate size derivatives, the term $-\partial_x[k(x)P(x, t)]$ represents the single-lipid association process, the term $-s(x)P(x, t)$ represents the aggregate division (splitting) process in which an aggregate of size x is split into two new aggregates, and the term $\int_x^\infty T_x(y)s(y)P(y, t) dy$ represents the addition of new aggregates of size x due to the division of aggregates of size larger than x . Here $T_x(y)$ is the probability density for generating an aggregate of size x (and one of size $y - x$) from an aggregate of size y . Finally, the term $-e(t)d(x)P(x, t)$ represents the outflow (or normalization) from the aggregates from a steady state chemostat. This term ensures that $P(x, t)$ is normalized to be a probability distribution.

Analytical steady state solutions for Equation 12 can be obtained, for example, with $d(x) = 1$ and k a constant, in which case the equilibrium aggregate size distribution is given by

$$P(x) = \frac{2}{\sqrt{k}} x \beta(x, k) e^{-\beta(x, k)x}, \quad \beta(x, k) = \frac{1}{\sqrt{k}} + \frac{x}{2k} \quad (13)$$

For details see [95]. Figure 11 displays the micellar size distribution in Equation 13.

The most comprehensive theoretical understanding of the detailed lipid self-assembly dynamics probably comes for a study of the processes in 3D bottom-up simulations. Depending on the relevant level of detail (time and length scales) for the questions we ask, we may use: (a) the molecular dynamics (MD) method for the processes on the small length and time scales, (b) the MD lattice gas method for the intermediate length

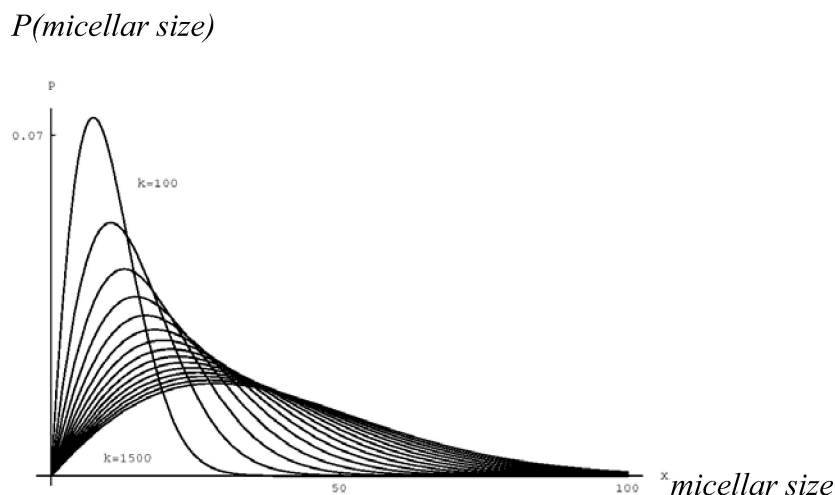


Figure 11. Equilibrium micellar size distribution from Equation 17 for values of k ranging from 100 to 1,500. Note how the choice of parameters yields small micelles with maximal probability of finding micelles of size $x_{\max} = \sqrt{k}(\sqrt{3} - 1)$ ($\partial_x P(x) = 0$). See text and [95].

and time scales, and (c) the lattice Boltzmann or Ginzberg-Landau continuous methods for the large length and time scales. In Section 5 a more detailed discussion of such a multilevel coupling of 3D simulations can be found. Perhaps the most appropriate simulation method for directly addressing molecular self-assembly processes is the MD lattice gas method, although it does not have enough predictive power as a standalone method.

The MD lattice gas simulation technique [52, 53, 66, 77] addresses molecular interactions at nanometer to micrometer length scales and over time scales up to milliseconds on workstations and seconds on supercomputers. This intermediate range suits them to model molecular self-organization and self-assembly processes involving ions, monomers, complex polymers, polymer aggregates (supermolecular structures), complex charged surfaces, and chemical reactions. The MD lattice gas method enables direct simulations of amphiphilic polymers self-assembling into micelles in aqueous environments (where the water molecules are simulated explicitly). The technique is a combination of traditional molecular dynamics and a lattice gas cellular automaton. Molecules (not atoms) are spatially confined to a three-dimensional lattice (i.e., both translational and rotational degrees of freedom are discrete), whereas the momentum space is continuous (particle velocity distributions are Maxwellian).

The translational movements are probabilistic and proportional to the velocities of the molecules. The rotational dynamics is an annealing process, which attempts to minimize the local potential energy. The collision operator locally preserves momentum, angular momentum, and kinetic energy. The particle interactions are represented by continuous fields, which are defined at discrete points between the molecular lattice points. During translational and rotational movement of the molecules, kinetic and potential energy is exchanged, but the total energy is conserved. The main advantage of using a discrete lattice is that the discreteness of the spatial dimensions numerically stabilizes the simulations and therefore enables simulations on time and length scales unreachable by traditional MD techniques.

Using a continuous momentum space together with nontrivial molecular interactions gives a model rich enough to be thermodynamically interesting for molecular self-

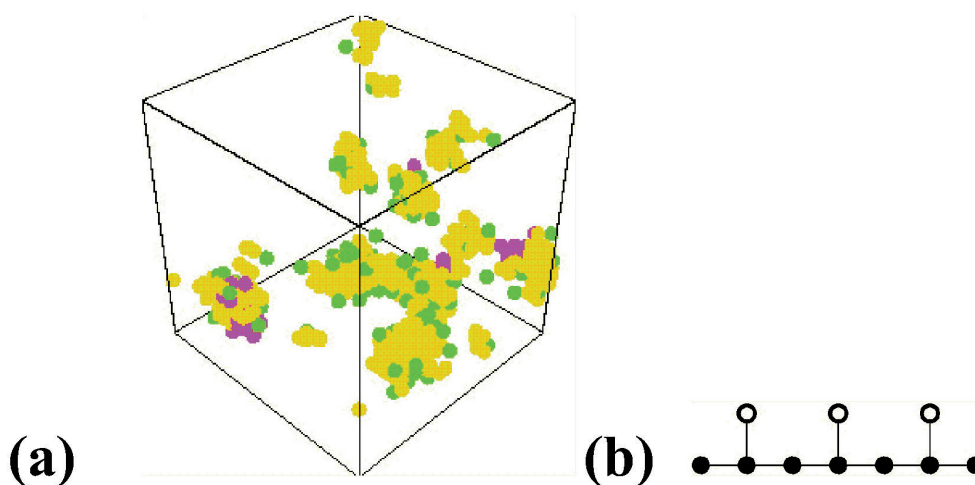


Figure 12. Micellar self-assembly and interactions with PNA-like polymers. Water is not shown. The yellow monomers are the hydrophobic tail parts of the lipids, and the green monomers are their hydrophilic head groups. The red polymers are short simplified PNA trimer strings [see (b)], which in the simulation are composed of a linear hydrophobic backbone with a side group on every other backbone element. Each PNA polymer has a hydrophobic backbone, which results in its affinity for the lipid aggregates. Each lattice cube is about 9 nm on each side, and typical simulation times for the situations shown are about a millisecond. Note that different lipid molecules probably should be used so that larger micelles are produced naturally. Larger lipid structures are necessary because they are better able to host longer PNA strings and loading of precursor molecules.

assembly processes. Figure 12 depicts a preliminary MD lattice gas study of PNA partition and loading in micelles, which we discuss in Section 4.2.

4.2 Loading (Solubilization) of the Lipid Aggregates

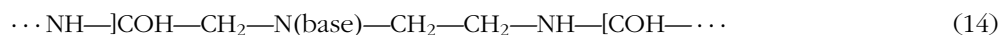
Thermodynamic driving as described above also occurs when hydrophobic polymers assemble in water; see Figure 10b. In this case less organized hydrophobic aggregates, or *blobs*, are formed, and (depending on the overall hydrophobicity of the polymers) a clear phase separation may occur, as seen for example for mineral oil in water. The oil-water picture may be a useful analogy for the lipid precursor assembly in water.

The thermodynamics of the loading of the lipid aggregates with precursors and templates is discussed in Figure 10b,c. As seen in Figure 10b, the precursor lipids, which are hydrophobic polymers, will most likely act similarly to mineral oil in water and phase-separate by forming large nonstructured hydrophobic clusters in the water. Adding amphiphilic lipid molecules, which are produced from the fragmentation of these precursor molecules, tends to break up the larger hydrophobic clusters into smaller aggregates and thereby form more interface area [58]. This occurs because the lipids act as surfactants, placing themselves with their hydrophobic tails in the precursor blobs while sticking their hydrophilic head groups into the water phase. In contrast, adding a small amount of all-hydrophobic precursor lipids to a solution of micelles in water tends to load the micelles with the precursor molecules, making the micelles bigger (swelling). Since the micelles now have a larger hydrophobic volume, they may be more rodlike when loaded; recall Equation 7.

Thus, as more surfactants (lipids) are being generated within the loaded micelles, as discussed in Figure 10c, the larger and more hydrophobic blobs will have a tendency to break up into smaller aggregates. In the end, if all precursor lipid is transformed into lipid, only micelles are left, and they will be of a particular size typical for the given surfactant and the physicochemical conditions. The important lesson to remember in

the context of proto-organisms is that the precursor-loaded micelles have a tendency to be larger than the structures formed by pure lipids. Thus there is a thermodynamic drive to divide the aggregates into smaller units once the metabolic processes have run their course and transformed all the (mineral-oil-like) precursor lipid into functional surfactant molecules. This aggregate division process of course has to be coordinated with the PNA replication process.

Also, the precursor PNA molecules will attach to the lipid aggregates—how strongly depends on the detailed composition of the PNA backbone. As a first approximation we assume that the base components of the PNA do not have a preference for the lipid phase (are not hydrophobic; they are identical to their DNA counterparts), so that we can assume that the main hydrophobic part of the PNA is within the backbone. Then we may approximate the free energy advantage for joining the lipid phase by adding the ΔG component contributions from the backbone. For each PNA monomer the backbone polymer piece is composed of a set of molecular groups (monomers) [91], where the brackets indicate the repeat of the structure:



The three CH_2 monomers are the only clearly hydrophobic groups of the backbone, so as a first approximation we may assume that the free energy of lipid attachment is less than $3 \Delta G_{\text{attach}}(\text{CH}_2)$, which is ~ -6 kJ/mol (recall the discussion of Equation 8), which is a weak free energy advantage. The free energy advantage of lipid attachment can be made much stronger if necessary, for example by the attachment of strongly hydrophobic peptide groups such as phenylalanine, valine, or leucine; recall the discussion in Section 3.3. Also, a direct attachment of the photosensitizer to the PNA backbone, (recall the discussion in Figure 6) will surely anchor the PNA backbone in the lipid phase. Preliminary MD lattice gas simulation studies of the PNA lipid aggregate (micelle) interactions are described in Figure 12.

In conclusion, it should be noted that at present we do not know the detailed overall dynamics of the complex aggregates as a function of the lipid concentrations, the loading of precursor molecules, and the PNA concentration. It is clear that self-assembly will occur due to the thermodynamic driving as described above, and we may also assume that a loaded micelle will become less stable as the lipid precursors within the aggregates are metabolized into functional surfactants (lipids). However, either direct experiments or detailed multilevel 3D simulation investigations have to be performed before we can make predictions about the dynamics of these complex aggregate systems.

4.3 Thermodynamics and Kinetics of the Proto-metabolism

Photofragmentation processes are found in both non-biotic and biotic systems, and technological applications for them exist [13, 14]. In Section 3 we discussed several implementations of photofragmentation processes that produce lipids and/or oligomers for the templating process, where the oligomers also act as photocatalysts. The overall energetic scheme of such a process is depicted in Figure 14, which depicts the photofragmentation of precursor lipid and precursor oligo molecules into functional lipids and oligomers. Since the fragmentation process involves a breakage of a covalent ester bond, the photon energy must be larger than this covalent bond energy. A typical free energy difference between the ground state and the excited state is 400 kJ/mol, and a typical free energy difference between the excited state and end products is 300 kJ/mol [62]. The actual energy levels as well as the kinetic reaction constants depend on the physicochemical details.

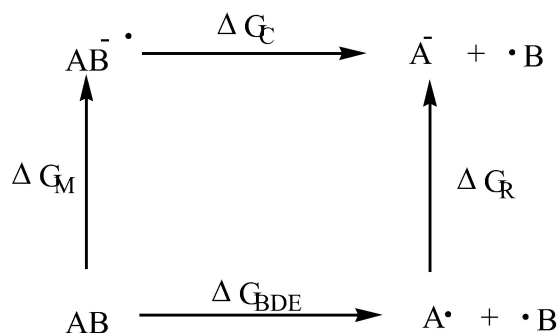


Figure 13. General scheme for the thermochemical cycles (or pathways) of the photofragmentation reaction. Starting in the lower left corner, two different pathways are in principle possible for the products in the upper right corner: direct breakage of the C–O bond in the neutral molecule AB to form two radicals A \cdot and \cdot B, followed by the one-electron reduction of the lipid to form the desired products A $^{\cdot-}$ and B $^\cdot$, or alternatively, a one-electron reduction of AB to form the radical AB $^{\cdot-}$ followed by the cleavage of the C–O bond in the charged radical. The upper pathway with an initial electron reduction of AB significantly weakens the C–O bond so that the fragmentation step happens more readily (and predictably), resulting in higher yields. Since both pathways result in the same products, the overall free energy difference for the two pathways is obviously the same. See text for details.

Photoinduced fragmentation reactions have been studied by our team as well as by other groups [13, 14, 71]. The key issue for a successful and high yield photofragmentation reaction is to bring the reactant into a state where the bond to be broken is significantly weakened. The thermodynamic properties and quantum efficiencies of these reactions depend upon the molecular structure of the reactants in addition to the reaction conditions. For the fragmentation reaction of phenacyl esters (PhCOCH₂–OCOR) as mentioned in Section 3, which can undergo photoinduced C–O bond scission in the presence of appropriate electron donor molecules, it is suggested that the conformation of the anion radical intermediates may play an important role in determining the rate of the bond fragmentation, and consequently the quantum efficiency of the reaction. The reactivity of the anion radical can be predicted based on the thermochemical cycles shown in Figure 13.

Since the free energy change is independent of pathway, the free energy change for the anion radical cleavage (ΔG_C) can be estimated from Equation 15, where ΔG_{BDE} represents the bond dissociation free energy of the neutral molecule, ΔG_R the free energy required for one-electron reduction of the radical fragment, and ΔG_M the free energy of one-electron reduction of the neutral molecule:

$$\Delta G_C = \Delta G_{BDE} + \Delta G_R - \Delta G_M. \quad (15)$$

The facile fragmentation (ΔG_C) of the photogenerated anion radical upon one-electron reduction (ΔG_M) can then be understood from the thermochemical cycle depicted in Figure 13: In general, the electron affinity (or reduction potential) of a neutral compound of starting material is always higher than that of the corresponding fragment radical ($\Delta G_R < \Delta G_M$). Therefore there is frequently a sizable difference in redox potentials, or electron affinity, between the neutral starting material and the fragments produced by homogeneous cleavage of the C–O bond (ΔG_{BDE}) on the one hand, and the starting material and the results of cleavage of the C–O bond in the anion radical (ΔG_C) on the other hand, which leads to the prediction of greatly reduced C–O bond energies in the anion radicals compared to the neutral molecule. As a consequence, the dissociation energies of specific bonds will be drastically lowered in the anion radicals compared to the corresponding neutral compounds.

To be more quantitative, the difference in free energy of reduction ($\Delta G_R - \Delta G_M$) can be shown⁸ to be proportional (with some factor F) to the difference in reduction potential of the radical and the neutral molecule [$E_{\frac{1}{2}}^{\text{ER}}(\text{A}^\cdot)$ and $E_{\frac{1}{2}}^{\text{RE}}(\text{AB})$]. Thus, the free energy change for anion cleavage (Equation 15) can be rewritten as Equation 16 below, where F is the previously mentioned proportionality factor. To evaluate the process by experimentally accessible observables such as the enthalpy (ΔH), Equation 16 can be transformed into Equation 17 using Equation 1:

$$\Delta G_C = \Delta G_{\text{BDE}} - F[E_{\frac{1}{2}}^{\text{ER}}(\text{A}^\cdot) - E_{\frac{1}{2}}^{\text{RE}}(\text{AB})] \quad (16)$$

$$\Delta H_C = \Delta H_{\text{BDE}} - F[E_{\frac{1}{2}}^{\text{ER}}(\text{A}) - E_{\frac{1}{2}}^{\text{RE}}(\text{AB})] - T[\Delta S(\text{AB}) - \Delta S(\text{AB}^{\cdot-})] \quad (17)$$

The last term in Equation 17 represents the difference in entropy change for the cleavage of the neutral molecule and the corresponding radical anion. If it is assumed that for these large delocalized species the solvation of the radical anion is similar to the solvation of the carboanion fragment formed upon cleavage, and if it is further assumed that the solvation entropies of the neutral molecule and the radical fragment are similar to each other (or are small compared to the entropies of the charged species), then this entropy term will be very small, so that $\Delta S(\text{AB}) = \Delta S(\text{AB}^{\cdot-})$, and thus the enthalpy for the bond cleavage can then be estimated from

$$\Delta H_C = \Delta H_{\text{BDE}} - F[E_{\frac{1}{2}}^{\text{RE}}(\text{A}^\cdot) - E_{\frac{1}{2}}^{\text{RE}}(\text{AB})] \quad (18)$$

Relating the bond dissociation enthalpy to the rate of cleavage of the radical anion requires the further assumption that there is no additional activation energy for this process, that is, there is no activation energy for the reverse reaction of the radical fragment with the carboanion. While this may be true for the reaction in the gas phase, in solution there may be some activation energy associated with solvent reorganization, particularly in polar solvents. Nevertheless, Equation 18 should be usable as the first approximation to predict the reactivity of the anion radical. For example, the redox potential for the reduction of phenacyl ester ($\text{PhCOCH}_2\text{--OCOR}$) is estimated to be -0.7 V versus standard calomel electrodes (SCE), and the redox potential for the reduction of radical PhCOCH_2^\cdot is estimated to be greater than 1.3 V versus SCE. Based on Equation 18, the bond cleavage energy for the phenacyl ester radical anion will be much less than that for the phenacyl ester, so that one should observe rapid fragmentation reactions of the photogenerated radical ion.

As a complement to the above thermodynamic calculations, detailed quantum-chemical calculations (ab initio) could also be applied, for example, to derive the bond energies for the different excitation stages of the photo-driven processes and to provide key parameters in the higher-level simulations [2, 85, 25, 92, 93]. Quantum-chemical calculations are currently feasible for finite molecular systems up to hundreds of atoms in size, although they are excessively computationally expensive. Such calculations could serve as a basic tool for understanding the main physical and chemical processes in small molecular systems and feed calculated input to the coarser models as well as to the experiments and the synthesis.

Quite a bit is known about the experimental and theoretical properties of charge transfer in the DNA double helix under different conditions [10, 69, 106, 40, 43]. Based on this body of work, we can make well-informed guesses about the charge transfer properties of PNA.⁹ We expect that charge transfer is faster in PNA than in DNA, par-

⁸ Values for $E_{\frac{1}{2}}^{\text{RE}}$ are tabulated and can be found in standard tables [62].

⁹ Personal communication with Alexander Burin, George Kalosakas, and Kim Rasmussen.

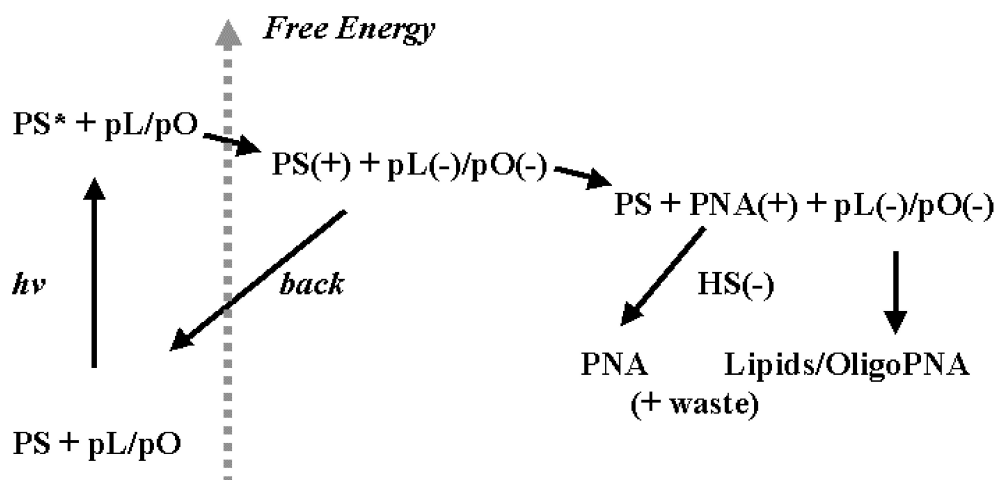


Figure 14. The thermodynamics of the proposed photo-driven proto-metabolism: The vertical axis indicates the relative free energy of the components. Only the overall reactions are depicted, not each detailed reaction step. From the bottom left: PS, pL, pO, and PNA denote a sensitizer, precursor lipid, precursor (PNA) oligomer, and PNA, respectively. Approximate values for the reaction constants for each main reaction are discussed in the text. As light energy ($h\nu$) is pumped into the system, the sensitizer is initially energized. This energy causes a charge separation between the sensitizer and precursor molecules (either pL or pO). This charge separation is very short-lived (because of the back reaction), unless a nearby electron relay system is present. Adenine within PNA is the immediate electron donor for the now activated sensitizer molecule, and adenine immediately thereafter gets an electron from guanine, which eventually will harvest an electron from a final electron donor as sulfite (HS^-). The resulting energized and negatively charged precursor molecule then fragments at the ester bond site, after which a functional lipid—or PNA oligomer—is created, together with waste. The overall reaction is an example of a photofragmentation reaction.

ticularly if the PNA duplex is in a lipid phase. The stronger the hybridization energies are, the colder the system is, and the fewer ions and water molecules we have near the PNA, the faster the charge transfer is expected to be. Charge transfer rates around 10^{12} s^{-1} are to be expected.

The key kinetic question for the proto-metabolic system is how to develop a set of processes with a good quantum yield as discussed above. This means that the balance between the back reaction (see Figure 14) and the electron charge transfer reaction is in favor of the charge transfer reaction. The typical reaction constant for a photoactivation process is $\sim 10^{-15} \text{ s}$, and the typical back reaction constant is $\sim 10^{-10} \text{ s}$. The charge transfer reaction constant thus needs to be somewhat smaller $\sim 10^{-11} \text{ s}$, which we also expect it to be from the above discussion. The rate-determining factor for the production of building materials is given by the efficiency of the charge-transfer chain. Thus, the presence of an electron relay system determines whether the yield of new lipid and PNA is high enough to counter the unavoidable degradation and diffusion processes.

Since there is no feedback with amplification (autocatalysis) either in the photosegmentation process or in the general redox process, the overall kinetics for the proto-metabolism is exponentially damped as a function of the driving (the energetic input). To demonstrate this important quality of the kinetics, and not to focus solely on photo-driven systems, we can for example assume a general water-membrane-water redox scheme for a vesicle as discussed in Figure 2. That gives us the following well-known simplified redox scheme:



where the input to the system is energy-rich molecules A and precursor molecules F. The waste products are B. The hydrophobic redox compound in the membrane exists in two oxidation stages C and D, which it is being cycled between. A acts as the external energy input to C (in the membrane), which is then transformed into D (also in the membrane), which in turn can modify some other compound F (precursor) into E (proto-organism building blocks). The solution to the kinetic equations derived from Equation 19 yields (under symmetric initial conditions assumptions and for $k_{AC} = k_{BF} = 1$)

$$\begin{aligned} A(t) &= F(t) = a_0 \exp(-c_0 t) \\ B(t) &= E(t) = b_0 + a_0[1 + \exp(-c_0 t)] \\ C(t) &= D(t) = c_0 \end{aligned} \quad (20)$$

where a_0 , b_0 , and c_0 are the initial concentrations of A, B, and C. As expected, it essentially is a smooth exponential decay toward thermodynamic equilibrium unless A is continuously added and B is continuously removed. As for the photo-driven reactions, the key redox kinetic question is to ensure high enough reaction rates k_{AC} and k_{BF} that the system produces enough building materials to sustain itself and grow.

In summary, the rate limiting steps for the proto-organism stemming from the metabolic processes will have either (a) too low yield in the production of functional oligomers, and/or (b) too low yield in the production of lipid molecules. In particular, if the proto-metabolic reactions are photochemical, it is important to note that (c) the reactions involved in the photokinetics are many, many orders of magnitude faster than the kinetics involved in the lipid aggregate (assembly, growth, and division) as well as the template-directed ligation and replication reactions (see the next subsection); (d) the proto-metabolic reactions will asymptotically reach a steady state production of the building blocks for the proto-organism if the system is continuously fed with the energy-rich electrons or photons and the appropriate precursor molecules.

4.4 Thermodynamics and Kinetics of the Proto-genes

The key thermodynamic questions for the proto-gene replication process are:

1. How tightly (expressed as free energy) is the PNA backbone associated with the lipid aggregate? We already discussed this question in Section 4.2.
2. How tightly associated are the PNA-PNA complexes in water at the water-lipid interface, and in the lipid phase?
3. What is the activation barrier for the template-directed amide bond formation in the membrane? Without template-directed ligation (polymerization), there is no replication.
4. Under which conditions is the PNA-PNA-lipid-water system in a balanced state with both hybridized and unhybridized PNA? Under which external conditions can we alter this balance towards disassociation without destroying the aggregate complex? Disassociated PNA-PNA duplexes are important for continued template reactions, which require single-stranded PNA.

These questions are discussed in Figure 15.

Melting curves for PNA duplexes indicate a high melting temperature of around 70° C even for PNA decamers (a higher melting temperature than that for the corresponding RNA or DNA decamers) [79]. To destabilize the PNA duplex it may be problematic just to raise the temperature, since we can easily lose the integrity of the lipid aggregate

at such high temperatures, although some surfactant aggregates have their most stable regime at around 60° C [81]. Experimentally it has been shown [91] that the free energy of base pairing (hydrogen bond formation in the PNA duplex) is about $\Delta G_{\text{base}} \approx -8$ kJ/mol per base pair or PNA monomer for PNA in water, which corresponds to a melting temperature of about 70° C at low salt concentration (the melting temperature decreases some with increasing salt concentration, but not very much). Thus, the PNA-PNA association is stronger than the estimated PNA-lipid association (ΔG_{lipid}) which is less than -6 kJ/mol per PNA monomer backbone segment. However, this balance may be significantly changed in the other direction by modifications of the PNA backbone; recall the discussion in Section 4.2. Also, the attachment of hydrophobic groups to the PNA backbone directly influences (decreases)¹⁰ the value of ΔG_{base} . As an example, let us assume that we attach a small hydrophobic side chain to the PNA backbone $[-\text{CH}-(\text{CH}_3)_2]$. As a first approximation we may assume that ΔG_{base} stays the same (~ -8 kJ/mol), but this side chain will approximately double the hydrophobicity of the backbone to about -12 kJ/mol per PNA monomer. Shifting this balance may also create a hydrophobic torque from each of the side chains that weakens the base pairing and thereby generates a metastable PNA complex.

The stability of PNA-PNA duplexes in organic solvents and in liposomal (micellar) solutions is not yet known. PNA-PNA duplex formation in aqueous solution is enthalpically driven with a large entropy loss [80]. Such duplexes are stabilized by a combination of hydrophobic, base stacking interactions and hydrogen bonding between the base pairs. We believe that the hydrophobic contribution dominates in water, as the nucleobases to a large extent exchange the hydrogen bonding to water molecules with hydrogen bonding to the complementary nucleobase. In contrast, it would be predicted that the situation should be opposite in organic solvents, where the formation of hydrogen bonds should provide a significant energetic stabilization, whereas base stacking should contribute less energy. Preliminary data in water-organic-solvent mixtures indicates a decreased PNA-PNA duplex stability in non-aqueous solvents.¹⁰ Thus the thermodynamics of these systems will eventually have to be determined by thermal denaturation (melting) experiments. Furthermore, it is expected that PNA duplexes will be more soluble in lipids than the single-stranded PNAs, due to shielding of the hydrophilic, hydrogen-binding faces of the nucleobases. These are all excellent questions to address with detailed MD simulation combined with experiments.

At present we do not know the exact activation energy associated with a lipid assisted amide bond formation in the template-directed PNA ligation, but lipid assisted amide bond formation from free peptides has been experimentally demonstrated by Luisi and coworkers [5] with reasonable kinetic yields (products in 24 h). These experiments also demonstrated a certain difference in the polymerization yields, depending on the detailed peptide chains. It is not unreasonable to expect a higher yield in template-directed amide bond formation, since the free energy barrier associated with the amide bond formation in the juxtaposed situation of a template-directed ligation should be lower than it is for amide bond formation between free polypeptides. In addition, recall (Section 3.3) that simple catalysts or changes in the external conditions could also enhance the ligation kinetics. Experiments will eventually be able to settle the PNA ligation question.

Assuming the above thermodynamic issues have been settled, a single-stranded PNA template C is located at the lipid-water interface, exposing the hydrophilic part of the bases to the water, while the hydrophobic backbone is sunk into the lipid layer. The system is fed with oligos A and B from the aqueous phase (or from oligos already attached to the lipid aggregate), so initially only partial double strands (AB or AC)

¹⁰ Peter Nielsen, private communication.

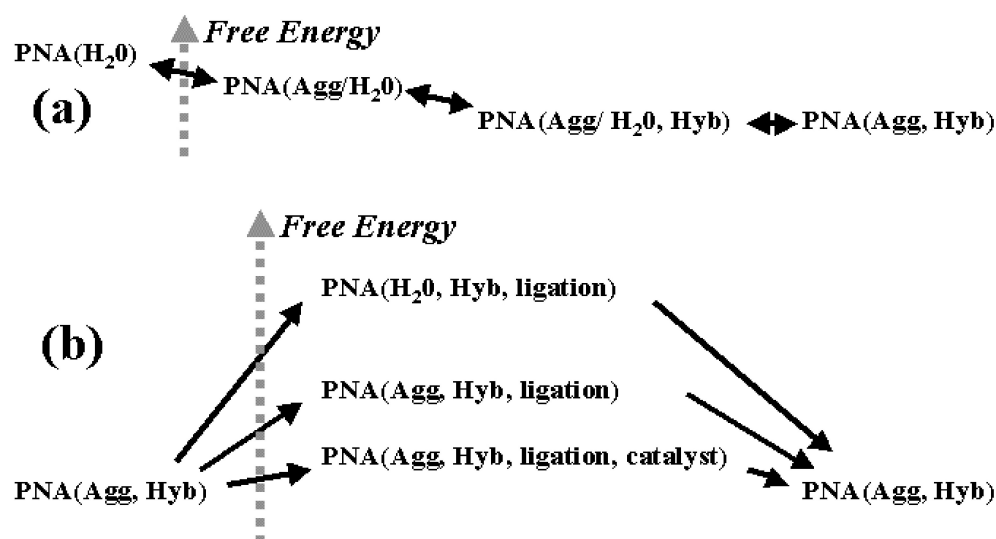


Figure 15. Thermodynamics of the lipid catalyzed PNA replication scheme. (a) PNA in water will attach to the lipid aggregate due to its hydrophobic backbone. If complementary strings meet, they will hybridize to form duplexes. Neutral diffusion of the duplex between the lipid aggregate surface and the interior of the aggregate can occur because the exterior of the duplex is hydrophobic. Since each step is thermodynamically downhill (or neutral) from left to right, it poses a problem to dehybirdize the duplex after a ligation has occurred, which is necessary for successive templating processes. A balance has to exist between the effective duplex hybridization energy and the energy gain from the exclusion of the hydrophobic backbone from the water phase. (b) Once the ABC duplex is formed, a condensation reaction has to occur between A and B to form C'. The activation energy in water is very high, but it is significantly lowered in the lipid aggregate. Perhaps the introduction of simple organic catalysts could lower the amide bond activation barrier even further.

are obtained by specific hydrogen bond formation. Once a full double strand ABC is formed, the whole surface of the complex is hydrophobic, so it will tend to sink further into the lipid phase. Here the ligation step can occur, which requires peptide bond formation. The ligation results in one plus and one minus template CC'. A small fraction of CC' will diffuse back to the lipid-water interface, where it may disassociate into its two strands, perhaps as the system goes through some cycle in external salt concentration or temperature. The balance between the hybridized CC' complex and the free C and C' strands depends on the hybridization energies (which depend on the PNA backbone properties), the temperature, the pH and the salt concentration of the aqueous phase, and several other factors. The important ingredient in our reasoning is that both C and C' remain anchored at the surface of the lipid aggregate, so that the proto-genes are permanently associated with the lipid aggregate. The kinetics of the replication is composed of two templating processes, followed by a template directed ligation step, with a subsequent disassociation, which can be expressed as



where all reactions are reversible except for the ligation. A detailed discussion of the dynamics of this process can be found in [78], where we use the methods developed in [104], [88], and [8]. Perhaps the most interesting finding is that the kinetics supports coexistence of multiple proto-genes due to an overall parabolic growth rate for the

template directed replication process,

$$c(t) \propto t^2 \quad (22)$$

where c is the total concentration of template in all its associations from Equation 21. The parabolic growth dynamics is generated from the interaction between a strict replication process and a product inhibition process. Note that this growth law is ruled out by the original quasispecies model developed by Eigen [20]. The quasispecies model does not allow for coexistence beyond the mutant cloud of the master sequence in homogenous solution. The possibility of fragmented genomes in principle opens up a much richer set of scenarios for further evolution. However, the template directed gene replication at the lipid interface, of course, does not bypass the limits on the amount of heritable information posed by the error threshold [20]. To address the overall consequences of this growth law, a more complete reaction kinetics study is needed, which integrates the details of the lipid aggregate growth and reproduction kinetics as well as the metabolic kinetics.

To summarize, two key experimental questions are still open with respect to proto-gene self-replication:

1. Although Luisi and coworkers have demonstrated peptide oligomer polymerization catalyzed by a lipid aggregate [5], the same process has not yet been demonstrated for PNA-like molecules. The thermodynamic problem associated with this process is discussed in Figure 14b, where the free energy barrier of activation is depicted.
2. Another open question is how to accomplish the necessary dehybridization of the double-stranded complex CC' , to allow a new templating process to occur. It is not yet clear under which conditions an appropriate force hierarchy can be established to form metastable PNA duplexes. The rate limiting steps in this lipid assisted PNA replication scheme are still to be determined experimentally, but we strongly suspect that the ligation and the dehybridization processes constitute the two most problematic (and slowest) reaction steps.

4.5 Thermodynamics and Kinetics of the Full System Replication: Life Cycle of the Proto-Organism

To complete the full life cycle of the proto-organism, defined as the generation of one new copy of itself, requires (a) the production of enough lipid to bud off a new aggregate, coordinated with (b) the production of (at least two) new oligomers that are ligated to form (one or more) minus templates, and finally (c) a partition of the templates in which each aggregate has at least one template.

In the following we need to know the proportions of the proto-organism. The weight of the proto-organism is made up of the approximately 100 lipid molecules, the PNA 10-mer, and one or more photosensitizer molecules, which comes to about $100 \times (200 \text{ g/mol})$ for the lipids (≈ 12 -carbon molecules), plus $10 \times (192 \text{ g/mol per PNA monomer})$, plus (say) $5 \times (250 \text{ g/mol})$ for the photosensitizer molecules (twice that much if each PNA monomer has a sensitizer molecule attached), which is approximately 23,000 g/mol. Thus, a single proto-organism weighs $(23 \times 10^3 \text{ g/mol}) / (6 \times 10^{23} \text{ mol}^{-1}) \approx 4 \times 10^{-20} \text{ g}$, which means that it is $\sim 10^7$ times smaller than the smallest known unicellular organism (at $\sim 0.1 \times 10^{-12} \text{ g}$). With a density close to that of water (1 g/ml), the proto-organism volume is about $4 \times 10^{-20} \text{ ml}$, corresponding to a diameter of about 4.5 nm (10^{-9} m). In summary: the proto-organism is much smaller and much simpler than actual unicells.

Although we do not know at this point how to experimentally orchestrate the proto-organism's life cycle, it should be clear from the calculations in the previous subsections (4.1–4.4) that multiple conditions should enable such an orchestration. Our picture of the full proto-organism life cycle is given in Figure 16.

The experimental conditions could, for example, be controlled so that the photo-driven PNA oligomer (a 3- to 5-mer) generation from PNA precursors occurs first, followed by ligation directed by the PNA template (a 6- to 10-mer). It is assumed that two or more PNA 6- to 10-mer strands can stay attached to the same aggregate. Now, having generated new (negative) proto-gene templates, precursor lipids could be released into the solution, which would result in a loading of precursors into the aggregates. The lipid aggregates swell up as the mineral-oil-like precursor lipids are incorporated. As the photo-driven generation of new lipid molecules creates more molecules that function as surfactants, the larger aggregate becomes unstable and splits into two new stable aggregates. If several PNA templates (requiring balance between the single-stranded PNA and the duplexes) are attached to the original aggregate, a certain (random) partition between the two lipid aggregates will occur. If we assume that the rate limiting step is the template-directed ligation and the subsequent dehybridization process, we have reason to believe that the full life cycle could occur in 24 h or less, just based on the kinetics of the amide bond formation process reported by Luisi and coworkers [5].

The kinetic relation between the PNA sequence and the photometabolic processes determines the evolutionary properties of the proto-organism. Recall that the metabolic kinetics is enhanced by a certain PNA sequence (GAGA...), which is the basis for an error prone, coupled aggregate–proto-gene selection, defining the elements of a Darwinian evolutionary process.

Evolution is one of only a few known universal principles of living systems. However, recently theory has been developed by West et al. [101, 102] to explain the universal nature of allometric scaling laws relating mass to other actual biological observables. Modern biology satisfies remarkably simple empirical scaling relations, for example, between the mass of an organism and the power of its metabolism. The proto-organism, due to its minimalist design, provides a possible origo for this relationship.

Using the proto-organism's mass and estimated metabolic rate, we obtain a power of about 5×10^{-22} W for the 4×10^{-20} g system, which is about two orders of magnitude below the expected metabolic power limit, if the observed scaling relation is extrapolated down to the proto-organism size scale. There also seems to be a universal energy efficiency for the life cycle of real-life unicells, which is independent of cell size ($\sim 0.1 \times 10^{-12}$ to $\sim 10^{-6}$ g), cell organization (prokaryotic vs. eukaryotic), and cell metabolism (aerobic vs. anaerobic), and it comes out to about [105]

$$\sim 500 \text{ J/g} \quad \text{compared to} \quad \sim 1,200 \text{ J/g} \quad (23)$$

based on the life cycle of the proposed proto-organism. The 1,200 J/g biomass comes from about 100 new lipid molecules and about two new PNA oligomers, which are able to polymerize (ligate) into a new template, within its life cycle. Each new lipid molecule costs about 300 kJ/mol to produce, and it weighs about 2000 g/mol, which yields about 1,500 J/g, which will be adjusted down a little if we also allow for the generation of the new template with associated photosensitizer, yielding $\sim 1,200$ J/g. Thus, the proto-organism has about half the energy-to-biomass efficiency of an actual unicell.

Two significant conditions, with opposite effect, have to be emphasized when making a comparison between an actual cell and the proposed protocell:

1. The proto-organism is on life support in contrast with the environment most actual cells live in. Our proto-organism is provided with precursor molecules, which are

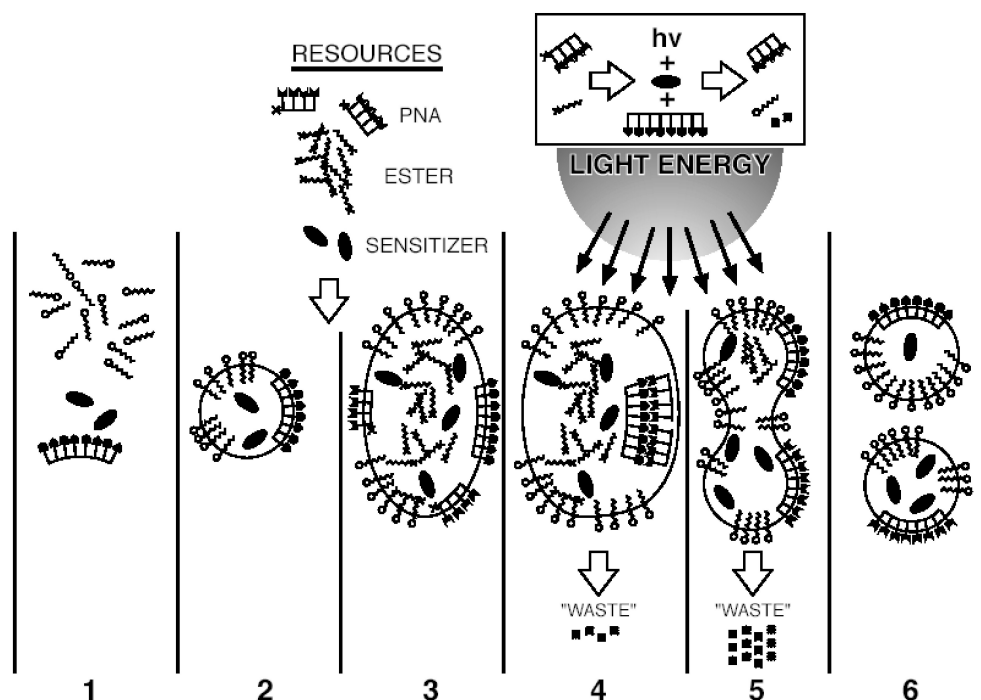


Figure 16. Full life cycle of the proposed proto-organism. 1, 2: Self-assembly of the proto-organism components—the lipids, the PNA, and the sensitizer molecules (Sections 3.1, 3.2, and 4.1). 3: Feeding the proto-organism with PNA and lipid precursors and more sensitizer molecules. The proto-organism swells up in particular as it is loaded with the many precursor lipids (Sections 3.1, 3.2, and 4.2). 4: As light is provided, the phenyl group is first fragmented from the precursor PNA, forming PNA oligomers that can ligate, once they are aligned by the template (Sections 3.1, 3.2, 3.3, and 4.3). 5: As the precursor lipids are turned into lipids (surfactants), the large aggregate becomes unstable and starts breaking up. A thermodynamic balance between hybridized and non-hybridized (double- and single-stranded) PNA has to be established to allow a reasonable partition of the proto-genes between the two aggregates (Sections 3.3 and 4.4). 6: The life cycle is complete, as the proto-organism has generated a copy of itself. The overall rate limiting steps are believed to be the PNA template directed ligation process and the balanced lipid-PNA and PNA hybridization kinetics. See text for more details.

quite similar in makeup to its molecular building blocks. This condition, however, should yield a *higher* proto-organism energy-biomass efficiency than for an actual cell.

2. Actual cells have evolved into highly efficient synthetic production units. On one hand, they are able to utilize resources of a much lower grade than their building blocks, and on the other hand, their metabolic processes operate much closer to the reversible regime than in the proto-organism, which is key to addressing the energy efficiency. An aerobic cell is, for example, able to produce up to 36 ATP molecules for each ingested sugar molecule.¹¹ These 36 ATP molecules can drive the synthesis of multiple high-grade biomolecules [46].

From these two significant conditions, we may define two different useful “distances” or metrics from the above energy efficiency discussion. The first metric may be defined as the (free) energetic distance between low-grade inorganics such as CO_2 and NH_4 , which autolithotrophs¹² can utilize, and the high-grade organics such as the provided

¹¹ A sugar molecule should also be considered as a high-grade molecule in this context.

¹² Microorganisms that can exist without any organic input, either as building blocks or as energy sources. These organisms are currently the true primary producers in our biosphere.

precursor lipids (phenacyl ester) for the proto-organism. This distance defines the degree to which a novel life form is on life support compared to an autolithotroph.

The second metric may be defined as the (free) energy distance to an ideal reversible metabolic conversion of the energy input all the way through the free energy utilized to biosynthesis. Modern life has evolved enzymes, which are used to lower the activation energies both in ATP production and in the ATP-driven synthesis, thereby lowering the required free energy for each of the metabolic steps. This metric defines the overall free energy efficiency of the metabolism. Using the first (life support) metric, it's clear that our proposed proto-organism does not address the question of how an autolithotrophic life form can emerge, which (for example) is the focus of Morowitz [61]. However, the first life forms may not have been autolithotrophs after all, as the primitive Earth might have had a much richer composition of complex organics than what we find today. If that is true, it seems reasonable to assume that the first life forms would have utilized the simplest possible metabolic processes and not at the onset developed more sophisticated and energy efficient metabolisms. Later life may have depleted this initial easy food, which would have forced more efficient metabolisms to evolve in response to this environmental change. We know that life has dramatically changed so many other environmental conditions over the history of the Earth (e.g., the high oxygen content of the atmosphere is biogenic), so why not also the initial global composition of free organic molecules?

The universal unit of contemporary life is a collection of physicochemical processes in a negative-entropic environment: a cell. The ability of a cell to generate negative entropy consists in its ability to utilize free energy and building blocks from the outer entropic environment and transform them into biomass. In this sense life is a natural free energy accumulating system. In the same manner, the proto-organism is a cluster of physicochemical processes in a negative-entropic environment (the lipid-PNA-sensitizer assembly).

Very importantly, it should be pointed out that in the above energetic calculations only the supplied external free energy used by the metabolic processes is included. The free energy associated with the self-assembly processes as they form the new aggregate is not taken into account. For the proto-organism the estimated free energy generated from self-assembly of the new aggregate is approximately $(100 \times 30 \text{ kJ/mol lipid}) + (150 \text{ kJ/mol PNA})$ (with photosensitizer), which yields about 3,150 kJ/mol. This has to be compared with about $(100 \times 300 \text{ kJ/mol}) + (600 \text{ kJ/mol}) = 30,600 \text{ kJ}$ for the metabolic production of the new building blocks. Thus, the self-assembly processes account for only about 10% of the total free energy change in one proto-organism life cycle, although no life as we know it would be possible without self-assembly. Also, in contemporary life it is the self-assembled organelles that keep all functionalities together and make life possible.

5 Discussion

In this section we summarize important issues already touched on in the previous technical sections as well as address potential future implications of our work. We provide a simple systemic solution to the question "How to generate a molecular proto-organism by rational design?" as we demonstrate how to integrate a container, a metabolism, and genes in a thermodynamically downhill manner. We provide experimental evidence for a subset of the processes, and we provide circumstantial and theoretical evidence for the remaining processes in terms of thermodynamic and kinetic arguments. However, experimental evidence for the integrated processes (the full proto-organism) does not yet exist.

5.1 What Is Known and What Is Not Known

Since it is just as important to emphasize what is still unknown as what is known about the proposed bridge between nonliving and living matter, we review the steps in Figure 16 and summarize the uncertainties. Process (1–2) is well understood, although the exact partitioning of PNA into the aggregates as a function of backbone details is not yet known. Process (2–3) is also well understood, although the exact loading capabilities of the different lipid aggregates are not yet known. Process (3–4) still has open questions, as it has not been demonstrated that PNA can act as a photocatalyst for both the lipid and the PNA precursors in praxis. Although every process is based on known chemistry, there is no guarantee that the composed system will work as the known chemistry predicts—even if all the subsystems work in isolation. For process (4–5) it is known that the production of more surfactants will destabilize the aggregates, but the appropriate conditions for template directed ligation of PNA in lipids with a subsequent metastable duplex state have not yet been demonstrated experimentally. Most of this article is devoted to an exploration of the underpinning issues associated with processes (3–4) and (4–5). Finally, process (5–6) follows from the generated lipid aggregate instability, although the partition of PNA templates (and photosensitizers) in the two daughter aggregates could in principle become an issue.

5.2 Origins of Life

It is hopefully clear that we do not claim to have addressed the historical question of the origins of life, although we do address the question of how to bootstrap a self-reproducing molecular machine, a “proto-organism,” from simple non-biological molecular species. We are seeking an example of the simplest and most fundamental organizational and structural bauplan of a proto-organism, not necessarily the earliest proto-organism.

We are focusing on a light-driven proto-metabolism, which may not have been the energy form utilized at the origins of life. Evidence elsewhere seems to indicate that the surface of the early Earth was too harsh an environment due to late accretion bombardments. The protected subsurface (the planetary crust) may very well have provided a richer chemistry with water, hydrocarbons, complex organics, minerals, and interfaces on the young Earth. In addition, phylogenetic studies based on RNA sequences seem to indicate that the redox-driven metabolism is older than the photo-driven metabolism. This all points to a subsurface redox-driven origin of life [15].

5.3 Artificial Life

It is argued in the previous sections that the proposed proto-organism mainly addresses open artificial life question 1, “Generate a molecular proto-organism in vitro,” but it also addresses important aspects of questions 4, “Simulate a unicellular organism over its entire life cycle,” and 5, “Explain how rules and symbols are generated from physical dynamics in living systems” [4]. Questions 4 and 5 are now discussed.

5.3.1 Question 4

It is clear that a stepwise approach is necessary to obtain a predictive simulation of the full life cycle of a modern unicell. It is also clear that a full virtual proto-organism could be constructed using one of several different simulation methods, or it could be constructed by means of a 3D multilevel simulation coupling different levels of description. We believe that the proposed proto-organism constitutes a useful stepping stone for the work leading up to a virtual unicell. The proposed proto-organism, together with other protocell proposals, could act as drivers and testbeds for part of the effort in systems biology that ultimately aims at developing whole cell simulations; see

for example [86]. Also, the virtual proto-organism constitutes a well-defined problem, which we hope that part of the artificial life community can attack, as it occurs at the boundary between nonliving and living matter as well as the boundary between life as it could be and life as we know it.

Traditionally, theoretical and computational studies of 3D molecular self-assembly systems have been performed with microscopic lattice models, Ginzburg-Landau theories, membrane theories, simple lattice gas models, and lately also MD simulations. While many of these models have been successful in explaining qualitative experimental results (mainly, producing topologically correct phase diagrams), only MD includes structural details of the interacting molecules: water, lipids, and other molecules. Many aspects of the dynamics of amphiphilic self-assembly depend strongly on the details in the structure of the molecules and their interactions. Water molecules form hydrogen-bond networks, the length of a lipid polymer determines the stability of assembled structures, the strength and geometric properties of the interaction between amphiphiles and water can directly affect the size distribution of micelles, and so on. However, MD is still too computationally expensive for large-scale (long time) simulations of self-assembly processes. The discrete spatial dimensions in the MD lattice gas technique numerically stabilize the time evolution and thereby enable simulations on time and length scales unreachable by traditional MD techniques. At the same time, using a continuous momentum space together with non-trivial molecule interactions gives a model rich enough to be thermodynamically interesting. Therefore it seems reasonable to couple (a) the MD method [31, 51, 90, 100] for small length and time scale processes, (b) the MD lattice gas method [53, 66, 77] for intermediate length and time scales, and (c) the lattice gas and lattice Boltzmann [7, 21, 41] or Ginzberg-Landau [38, 87] continuous methods for larger length and time scales simulating small proto-organism populations over many life cycles. Although the MD lattice gas as a standalone technique can address 3D molecular self-assembly processes, it should be noted that this method alone is not reliable enough to make precise experimental predictions. Only in concert with the more detailed MD simulations and experimental results from related situations as a reference can we expect to make true predictions about experimental systems at the complexity level of the proposed proto-organism.

5.3.2 Question 5

We may also discuss the emergence of rules and symbols in the proto-organism context. The emergence of a biological system from chemistry comes from the observed functionalities of the proto-organism, one of these functionalities being the presence of (proto-)genes, which can induce higher kinetic efficiency of the organism's metabolism. The nucleobases encode the metabolic efficiency in a very direct (physical) manner by acting as an electron relay chain. Nevertheless we contend that ours is a demonstration example right at the transition between physical dynamics and symbolic encoding dynamics, because of the nature of the relation between the functionality of the gene sequence and the way it is transmitted (encoded) to the next generation.

Continuing this way of thinking, we may be tempted to conjecture that "chemistry becomes biology" at the point where an organizational closure is established between physicochemical processes that naturally occur and interactions that tweak (enhance) their mutual yield. By this we may ask whether the initial biological organization just is a slight (later, as life evolves, it of course becomes a significant) control (tuning) on top of physicochemical processes that occur in any event. Our insight from molecular self-assembly processes and the effect of the gene sequences on metabolic reaction kinetics suggests this view.

5.4 Living Technology¹³ Applications

Development of proto-cells will be an accomplishment that derives value not from a singular application, but from proto-cells' ability to open the door to a huge array of applications. It will, if successful, be a threshold technological advance such as the silicon revolution (development of the transistor), the mapping of the human genome, and quantum computing. Pohorille and Deamer [70] recently made the case for the technological potential of proto-cells:

Artificial cells designed for specific applications offer unprecedented opportunities for biotechnology because they allow us to combine properties of biological systems such as nanoscale efficiency, self-organization, and adaptability for therapeutic and diagnostic applications . . . it will become possible to construct communities of artificial cells that can self-organize to perform different tasks and even evolve in response to changes in the environment.

Proto-cells have not been developed in the laboratory yet, because the necessary cell chemistry and architecture have previously been regarded as too complex for engineering. We believe the time is now ripe for their development, for the following reasons:

1. A novel and significantly simpler design has now been developed and presented in this article.
2. Most of the necessary chemical mechanisms have been independently implemented in vitro, and calculations indicate that a full system is feasible.
3. Combinatorial chemistry and automated laboratory techniques enable more efficient search of possible operation conditions for proto-cells.
4. Advances in the theory and simulation of self-assembly and molecular self-organization have reached the point that they can propel the technology to realize the proto-cell concept.
5. Last but not least, vast economic markets will become accessible as the proto-cell technology spans many application areas, including chemical, medical, environmental, structural, and computational technology. Proto-cells will be a key component of a vast array of different applications of living technology, all based on autonomy, robustness, self-repair, intelligence, and self-replication. It also seems necessary to resolve the fundamental problem of assembling a living system before key macroscopic (robotic) applications become available [9].

5.5 Ethical Concerns

Generating life de novo will elicit public reactions. The reactions will probably be along two lines: (a) environmental concerns that the life-producing technology could get out of control, and (b) religious and moral concerns, based on the beliefs that humankind must refrain from certain endeavors on grounds that they are fundamentally immoral.

Ethical issues related to the creation of cells other than natural cells have a long history, but were taken up in the press recently as fallout from announcement of the research on minimal cells at the Institute for Genomic Research [26, 36]. Their press

¹³ Concepts and perspectives about the emerging living technology and ethical concerns about it were developed by Mark Bedau, John McCaskill, Norman Packard, and Steen Rasmussen in their working group meeting in June 2001.

release and links to subsequent news stories may be found on their website.¹⁴ The resulting minimal cell was dubbed “Frankencell” in the non-technical press. Similar concerns about nanostructures capable of proliferating in natural environments have been formulated in the nanotechnology community [56]. However, genetically coded proliferation is not a risk in itself, which, for example, can be seen from its current ubiquitous use in molecular biology labs in the form of the polymerase chain reaction (PCR).

Regarding the environmental objections, the special control of proto-cell design and complementation, implicit in the approach proposed here, enables the designers to build in various forms of protection, such as dependence of the metabolism and the genetics on substances that do not occur naturally and must be provided to allow the cells’ continued existence and proliferation.

Regarding the religious and moral objections, reaction in the press directly after the Frankencell ethical panel report had no explicit religious overtones, but subsequent references to Frankencell in web discussions had extensive negative religious overtones. The ethical issues seem small in comparison with those presented by stem-cell and cloning research. In fact, proto-cells could provide an ethically cleaner route to providing subsets of functionality found in those research areas.

Finally we should also mention that the development of new carbon-chemistry-based life-forms could at least theoretically have biothreat application. In the long term, for example, it should be possible to engineer or evolve hitherto unknown human pathogens [30]. As with all technologies, the good and the evil possibilities arise together. We agree with the conclusions of the ethical panel reviewing Venter’s earlier Frankencell enterprise, which stated that if the ultimate goal was to benefit humankind and if all appropriate safeguards were followed, then the project could be regarded as ethical.

6 Conclusion

A simple, step-by-step, thermodynamically downhill assembly of a proto-organism is proposed. The resulting aggregate would be capable of feeding from the environment, growing and replicating, and dying under environmental stress, as well as undergoing Darwinian evolution. In short, the proto-organism would be alive at least with respect to the most common definition of the term. This is perhaps the first concrete proto-organism picture where all molecules and processes are explicitly given, and it is certainly much simpler than other recent proposals due to Luisi et al. [49], Szostack et al. [89], and Pohorille and Deamer [70]. Physicochemical implementations of the proto-organism are presented, and the associated experimental results and problems are discussed. Using micellar structures as a lipid proto-container with a photo-driven proto-metabolism, it is shown that it is possible to produce lipids as well as PNA oligomers from appropriate precursor molecules, and it is shown how the metabolic kinetics will depend on the encoded proto-gene sequences.

The thermodynamic and kinetic properties of the detailed subsystems are derived or simulated, and important conclusions can be drawn about the appropriate balance between key processes. Reaction kinetic analysis of the metabolism and the template-directed replications is presented. Also, reaction kinetics and detailed 3D simulations of the lipid self-assembly are presented.

A full life-cycle analysis of the proto-organism, partly based on experimental results and partly derived from calculations, shows that it will operate at about half the overall energy efficiency (~ 1.2 kJ/g biomass) of living unicells. The estimated generation time of the proto-organism is of the order of 24 h, and the proto-organism can be as small as 5 nm in diameter and weigh about 10^7 times less than the smallest unicell (diameter

¹⁴ www.tigr.com.

$\sim 0.5 \mu\text{m}$). It should be noted that the free energy associated with the molecular self-assembly processes only accounts for about 10% of the necessary metabolic energies. However, it is this 10% that enables life to exist, for this modest free energy enables the formation of the functional basis structures.

Finally, it should be reiterated that the minimalistic thermodynamic coupling between the container, the metabolism, and the genes could also be realized using a redox-based metabolism. It might even be possible to use RNA or some other charged template instead of PNA for the proto-genes by using oppositely charged lipid aggregates. We believe that the proposed proto-organism picture is quite general, although we are aware that multiple alternative transitions between nonliving and living matter are possible. Although several experimental pieces in the proto-organism puzzle are still missing, it is our hope that this detailed account of a particular bridge from nonliving to living matter can help focus more attention on this fascinating problem.

Acknowledgments

Development of the presented level of detail for the proposed proto-organism has taken several years, and it has involved help from many people. Special thanks to Klaus Lackner, Luigi Luisi, Peter Nielsen, Shelly Copley, David Deamer, David Whitten, Peter and Barbel Stadler, Daniel Yamins, Stirling Colgate, and Kolbjorn Tunstrom, who have all contributed either important ideas or technical details to this work. Over the last year Mark Bedau, John McCaskill, and Norman Packard have also provided important insight and perspectives to the proto-organism effort, in particular with respect to future proto-cell applications and the ethical context as well as its place in the broader context of living technology. In addition we want to thank Charles Apel, Bernd Mayer, Gottfried Kohler, William Newman, Michael O'Kelly, Donatella Pasqualinie, Andrew Pohorille, and Lau Sennels for valued discussions and support.

This work is supported in part by an LDRD grant from the Center for Space Science and Exploration (CSSE) and the LDRD grant entitled Computational Proto-organisms, both at Los Alamos National Laboratory.

References

1. Apel, C., Deamer, D., & Mautner, M. (2002). Self-assembled vesicles of monocarboxylic acids and alcohols: Conditions for stability and for the encapsulation of biopolymers. *Biochemical Biophysics Acta*, 1559, 1–9.
2. Alexander, A. J., & Zare, R. N. (2002). Molecular tennis—Flat smashed and wicked cuts. *Accounts of Chemical Research*, 33(4), 199–205.
3. Bagley, J. R., & Farmer, J. D. (1991). Spontaneous emergence of a metabolism. In C. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (Eds.), *Artificial life II* (pp. 93–140). Redwood City, CA: Addison-Wesley.
4. Bedau, M., McCaskill, S. J., Packard, N. H., Rasmussen, S., Adami, C., Green, D. G., Ikegami, T., Kaneko, K., & Ray, T. S. (2000). Open problems in artificial life. *Artificial Life*, 6, 363–376.
5. Blocher, M., Liu, D., Walde, P., & Luisi, P. L. (1999). Liposome assisted selective polycondensation of α -amino acids and peptides. *Macromolecules*, 32, 7332–7334.
6. Boerlijst, M., & Hogeweg, P. (1991). Self-structuring and selection: Spiral waves as a substrate for prebiotic evolution. In C. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen (Eds.), *Artificial life II*, SFI Vol. X (pp. 255–276), Redwood City, CA: Addison-Wesley.
7. Boghosian, B., Coveney, P., & Love, P. (2000). A three dimensional lattice gas model for amphiphilic fluid dynamics. *Proceedings of Royal Society of London*, 456A, 1431–1454.
8. Borghans, J. A. M., de Boer, R. J., & Segel, L. A. (1996). Extending the quasi-steady state approximation by changing variables. *Bulletin of Mathematical Biology*, 58, 43–63.

9. Brooks, R. (2001). The relationship between matter and life. *Nature*, 409, 409–411.
10. Bruinsma R., Gruner, G., D'Orsogna, M. R., & Rudnick, J. (2002). Fluctuation-facilitated charge migration along DNA. *Physical Review Letters*, 85, 4393–4396.
11. Böhler, C., Nielsen, P. E., & Orgel, L. E. (1995). Template switching between PNA and RNA oligonucleotides. *Nature*, 376, 578–581.
12. Cech, T. R. (1986). A model for RNA-catalyzed replication of RNA. *Proceedings of the National Academy of Science of the U.S.A.*, 83, 4360–4363.
13. Chen, L., & Whitten, D. (1995). Photoinduced electron transfer double fragmentation: An oxygen-mediated radical chain process in the cofragmentation of aminopinacol donors with organic halides. *Journal of the American Chemical Society*, 117, 6398–6399.
14. Chen, L., Lucia, L., & Whitten, D. (1998). Cooperative electron transfer fragmentation reactions: Amplification of a photoreaction through a tandem fragmentation acceptor and donor pinacols. *Journal of the American Chemical Society*, 120, 439–440.
15. Colgate, S., Rasmussen, S., Solem, J. C., & Lackner, K. (2003). An entropy and astrophysically based strategy for understanding a possible universal origin of life. *Advances in Complex Systems*, in press.
16. Deamer, D. W., & Pashley, R. M. (1989). Amphiphilic components of the Murchison carbonaceous chondrite: Surface properties and membrane formation. *Origins of Life and Evolution of the Biosphere*, 19, 21–38.
17. Deamer, D. W. (1992). Polycyclic aromatic hydrocarbons: Primitive pigment systems in the prebiotic environment. *Advances in Space Research*, 12, 183–189.
18. Deamer, D. W. (1997). The first living systems: A bioenergetic perspective. *Microbiology and Molecular Biology Reviews*, 61, 239–262.
19. Deamer, D. W. (1998). Membrane compartments in prebiotic evolution. In A. Brack (Ed.), *The Molecular Origins of Life* (Chap. 8, pp. 189–205). Cambridge, UK: Cambridge University Press.
20. Eigen, M. (1971). Self-organization of matter and the evolution of macromolecules. *Naturwissenschaften*, 58, 465–523.
21. Evans, D. F., & Ninham, B. W. (1986). Molecular forces in the self-organization of amphiphiles. *Journal of Physical Chemistry*, 90, 226–234.
22. Falvey, D., & Banerjee, A. (1997). Protecting groups that can be removed through photochemical electron transfer: Mechanistic and product studies on photosensitized release of carboxylates from phenacyl esters. *Journal of Organic Chemistry*, 62, 6245–6251.
23. Farmer, J. D., Kauffman S., & Packard, N. H. (1986). Autocatalytic replication of polymers. *Physica D*, 22, 50–67.
24. Ferris, J. P., Hill, Aubrey, R., Jr., Liu, R., & Orgel, L. E. (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381, 59–61.
25. Foresman J. B., Headgordon, M., Pople, J. A., & Frisch, M. J. (1992). Toward a systematic molecular-orbital theory for excited states. *Journal of Physical Chemistry*, 96, 135–149.
26. Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Furhmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J.-F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Scott Peterson, N., Smith, H. O., Hutchison, C. A., III, & Venter, J. C. The minimal gene component of *Mycoplasma genitalium*. *Science*, 270, 397–403.
27. Ganti, T. (1997). Biogenesis itself. *Journal of Theoretical Biology*, 187, 583–593.
28. Gestland, R., Cech, T., & Atkins, J. (1999). *The RNA world*. Cold Spring Harbor, MA: Cold Spring Harbor.
29. Gilbert, W. (1986). The RNA world. *Nature*, 319, 618.

30. Gillis, J. (2002). Scientist planning to make new form of life. *Washington Post*, Nov. 21, A01.
31. Goetz, R., Gompper, G., & Lipowsky, R. (1999). Mobility and elasticity of self-assembled membranes. *Physics Review Letters*, 82(1), 221–224.
32. Greiner N. R., Rogers Y., & Spall, D. (1990). *Chemistry of detonation soot: More diamonds and volatiles* (Los Alamos Report LA-11837-MS). See also: Badziag, P., Verwoerd, W. S., Ellis, W. P., & Greiner, N. R. (1990). Nanometer-sized diamonds are more stable than graphite. *Nature*, 343, 244.
33. Haaïma, G., Lohse, A., Buchardt, O., & Nielsen, P. E. (1996). Peptide nucleic acids (PNAs) containing thymine monomers derived from chiral amino acids: Hybridization and solubility properties of D-lysine PNA. *Angewandte Chemie, International Edition in English*, 35, 1939–1941.
34. Hargreaves, W. R., & Deamer, D. (1978). Liposomes from ionic, single-chain amphiphiles. *Biochemistry* 17, 3759–3768.
35. Hotani, H., Lahoz-Beltra, R., Combs, B., Hameroff, S., & Rasmussen, S. (1992). Liposomes, microtubules, and artificial cells. *Nanobiology*, 1(1), 61–74.
36. Hutchison, C. A., Peterson, S., Gill, S., Cline, R., White, O., Fraser, C., Smith, H., & Venter, C. (1999). Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, 286, 2165–2169.
37. Hyrup, B., Egholm, M., Nielsen, P. E., Wittung, P., Nordén, B., & Buchardt, O. (1994). Structure-activity studies of the binding of modified peptide nucleic acids (PNAs) to DNA. *Journal of the American Chemical Society*, 116, 7964–7970.
38. Jiang, Y., Lookman, T., & Saxena, A. (2000). Phase separation and shape deformation of two-phase membranes. *Physical Review E*, 61, R57.
39. Jonston, W., Unrau, P., Lawrence, M., Glasner, M., & Bartel, D. (2001). RNA-catalyzed RNA polymerization: Accurate and general RNA-template primer extension. *Science*, 292, 1319–1325.
40. Kalosakas, G., Rasmussen, K. O., & Bishop, A. R. (2003). Charge trapping in DNA due to intrinsic vibrational hot spots. *Journal of Chemical Physics*, 118(8), 3731–3737.
41. Kang, Q., Zhang, D., Chen, S., & He, X. (2002). Lattice Boltzmann simulation of chemical dissolution in porous media. *Physical Review E*, 65, 036318.
42. Kauffman, S. (1986). Autocatalytic sets of proteins. *Journal of Theoretical Biology*, 119, 1–24.
43. Komineas, S., Kalosakas, G., & Bishop, A. R. (2002). Effects of intrinsic base pair fluctuations on charge transfer in DNA. *Physical Review E*, 65(6), pt. 1, 061905.
44. Langton, C. G. (1984). Self-reproduction in cellular automata. *Physica D*, 10, 135–144.
45. Lee, D., Granja, J., Martinez, J., Severin, K., & Ghadiri, M. (1996). A self-replicating peptide. *Nature*, 382, 525–528.
46. Lengeler, J. W., Drews, G., & Schlegel, H. G. (1999). *Biology of the prokaryotes*. Cambridge, MA: Blackwell Science.
47. Luisi, P. L., Giomini, M., Pileni, M., & Robinson, B. (1988). Reverse micelles as hosts for proteins and small molecules. *Biochimica et Biophysica Acta*, 947, 209–246.
48. Luisi, P. L., Bachmann, P. A., & Lang, J. (1992). *Nature*, 357, 57.
49. Luisi, P. L., Walde, P., & Oberholzer, T. (1994). Enzymatic synthesis in self-reproducing vesicles: An approach to the construction of a minimal cell. *Berichte der Bunsengesellschaft für Physikalische Chemie*, 98, 1160–1165.
50. Luisi, P. L., Veronese, A., & Berclaz, N. (1998). Photoinduced formation of bilayer vesicles. *Journal of Physical Chemistry B*, 102(37), 7078–7080.
51. Maillet J.-B., Lachet, V., & Coveney, P. (1999). Large scale molecular dynamics simulation of self-assembly processes in short and long chain cationic surfactants. *Physical Chemistry Chemical Physics*, 1, 5277–5290.

52. Mayer, B., Koehler, G., & Rasmussen, S. (1997). Simulation and dynamics of entropy driven, molecular self-assembly processes. *Physical Review E*, 55, 1–11.
53. Mayer, B., & Rasmussen, S. (2000). Dynamics and simulation of self-reproducing micelles. *International Journal of Modern Physics C*, 11, 809–826.
54. McCaskill, J. S. (1993). Inhomogeneous molecular evolution. In *Jahresbericht 1992/1993*. Jena, Germany: Institut für Molekulare Biotechnologie IMB.
55. McCaskill, J. S. (1997). Spatially resolved in vitro molecular ecology. *Biophysical Chemistry*, 66, 145–158.
56. Merkle, R. (1992). The risks of nanotechnology. In B. Crandall & J. Lewis (Eds.), *Nanotechnology—Research and Perspectives* (pp. 287–294). Cambridge, MA: MIT Press.
57. Miller, S., & Urey, H. (1959). Organic compound synthesis on the primitive Earth. *Science*, 130, 245–251.
58. Mittal, K. L., & Lindman, B. (Eds.) (1984). *Surfactants in solution*. New York: Plenum.
59. Monnard, M., Apel, C., Kanavarioti, A., & Deamer, D. W. (2002). Influence of ionic inorganic solutes on self-assembly and polymerization processes related to early forms of life: Implications for a prebiotic aqueous medium. *Astrobiology*, 2, 139–152.
60. Morowitz, H., Deamer, D., & Heinz, B. (1988). The chemical logic of a minimal protocell. *Origins of Life and Evolution of the Biosphere*, 18, 281–287.
61. Morowitz, H. J. (1992). *The origins of cellular life: Metabolism recapitulates biogenesis*. New Haven, CT: Yale University Press.
62. Murov, S. L., Carmichael, I., & Hug, G. L. (1993). *Handbook of photochemistry*. New York: Marcel Dekker.
63. Nelson, K., Levy, M., & Miller, S. (2000). Peptide nucleic acid rather than RNA may have been the first genetic molecule. *Proceedings of the National Academy of Sciences of the U.S.A.*, 97, 3868–3871.
64. Nielsen, P. E. (1993). Peptide nucleic acid (PNA): A model structure for the primordial genetic material? *Origins of Life and Evolution of the Biosphere*, 23, 323–327.
65. Nielsen, P. E., Engholm, M., Berg, H., & Buchart, O. (1991). Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science*, 254, 1497–1500.
66. Nilsson, M., Rasmussen, S., Mayer, B., & Whitten, D. G. (2003). Constructive molecular dynamics lattice gases: 3-D molecular self-assembly. In D. Griffeth & C. Moore (Eds.), *New constructions in cellular automata* (pp. 275–290). New York: Oxford University Press.
67. Ono, N., & Ikegami, T. (1999). Model of self-replicating cell capable of self-maintenance. In D. Floreano et al. (Eds.), *Advances in artificial life* (pp. 399–406). Berlin: Springer-Verlag.
68. Oparin, A. I. (1924). *The origin of life on the Earth* (1st ed., Pabochii); 3rd ed. (1957). Oliver and Boyd.
69. Peyrard, M., & Bishop, A. R. (1989). Statistical mechanics of nonlinear model for DNA denaturation. *Physical Review Letters*, 62(23), 2755–2758.
70. Pohorille, A., & Deamer, D. (2002). Artificial cells: Prospects for biotechnology. *Trends in Biotechnology*, 20, 123.
71. Popielarz, R., & Arnold, D. R. (1990). Radical ions in photochemistry. Part 24. Carbon-carbon bond cleavage of radical cations in solution: Theory and application. *Journal of the American Chemical Society*, 112, 3068–3082.
72. Porter, R. (1973). Zinin reduction. *Organic Reactions*, 20, 455–481.
73. Püschl, A., Sforza, S., Haaima, G., Dahl, O., & Nielsen, P. E. (1998). Peptide nucleic acids (PNAs) with a functional backbone. *Tetrahedron Letters*, 39, 4707–4710.
74. Rajagopalan, R. (2001). Simulation of self-assembly systems. *Current Opinion in Colloid and Interface Science*, 6, 357–365.

75. Rasmussen, S. (1995). *Aspects of instabilities and self-organizing processes*. Ph.D. thesis, Department of Physics, Technical University of Denmark.
76. Rasmussen, S. (1988). Toward a quantitative theory of the origin of life. In C. Langton (Ed.), *Artificial life*. Redwood City, CA: Addison-Wesley.
77. Rasmussen, S., Baas, N. A., Mayer, B., Nilsson, M., & Olesen, M. W. (2001). Ansatz for dynamical hierarchies. *Artificial Life*, 7, 329–353.
78. Rasmussen, S., Chen, L., Stadler, B., & Stadler, P. (2003). Proto-organism kinetics: Evolutionary dynamics of lipid aggregates with genes and metabolism. *Origins of Life and Evolution of the Biosphere* (in press).
79. Ratilainen, T., Holmen, A., Tuite, E., Haaïma, G., Christensen, L., Nielsen, P. E., & Norden, B. (1998). Hybridization of peptide nucleic acid. *Biochemistry*, 37, 12331–12342.
80. Ratilainen, T., Holmén, A., Tuite, E., Nielsen, P. E., & Nordén, B. (2000). Thermodynamics of sequence-specific binding of PNA to DNA. *Biochemistry*, 39, 7781–7791.
81. Rosen, M. H. J. (1988). *Surfactants and interfacial phenomena*. New York: Wiley-Interscience.
82. Saghatelian, A., Yokobyashi, Y., Soltani, K., & Ghadiri, M. R. (2001). A chiroselective peptide replicator. *Nature*, 409, 797–401.
83. Sagre, D., Ben-Eli, D., & Lancet, D. (2000). Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proceedings of the National Academy of Sciences of the U.S.A.*, 97, 4112–4117.
84. Sayama, H. (1998). Introduction of structural dissolution into Langton's self-replicating loops. In C. Adami, R. K. Belew, H. Kitano, & C. E. Taylor (Eds.), *Artificial life VI* (pp. 114–122). Cambridge, MA: MIT Press.
85. Schnell M., Muhlhauser M., & Peyerimhoff, S. D. (1992). Ab initio MRD-CI study of excited states of chloromethanol ClCH_2OH and photofragmentation along C–O and C–Cl cleavage. *Chemical Physics Letters*, 344, 519–526.
86. Schwehm, M. (2002). *Parallel stochastic simulation of whole-cell models* (Internal report). University of Tübingen.
87. Smit, B., Hilberts, P. A. J., Esselink, K., Rupert, L. A., & van Os, N. M. (1991). Simulation of oil/water/surfactant systems. *Journal of Physical Chemistry*, 95(16), 6361–6368.
88. Stadler, B., & Stadler, P. (2002). Molecular replicator dynamics (Preprint series 02-09-049). Santa Fe, NM: Santa Fe Institute.
89. Szostak, J., Bartell, D., & Luisi, P. L. (2001). Synthesizing life. *Nature*, 409, 387–390.
90. Tieleman, T., van der Spoel, D., & Brendsen, H. J. C. (2000). Molecular dynamics study of dodecylphosphochloride micelles at three different aggregate sizes. *Journal of Physical Chemistry B*, 104(27), 6380–6388.
91. Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P. E., Norden, B., & Graslund, A. (1996). Ionic effects on the stability and conformation of peptide nucleic acid complexes. *Journal of American Chemical Society*, 118, 5544–5552.
92. Tretiak, S., & Mukamel, S. (2002). Density matrix analysis and simulation of electronic excitations in conjugated and aggregated molecules. *Chemical Reviews*, 102(9), 3171–3212.
93. Tretiak S., Saxna A., Martin, R. L., & Bishop, A. R. (2002). Conformational dynamics of photoexcited conjugated molecules. *Physical Review Letters*, 89(9), 097402.
94. Tuck, A. (2002). The role of atmospheric aerosols in the origins of life. *Surveys In Geophysics*, 23, 379–409.
95. Tunstrom, K. (2002). *Mathematical toy models for micelle self-assembly*. M.S. thesis, Department of Mathematics, Trondheim NTNU and Los Alamos National Laboratory.
96. von Kiedrowski, G. (1986). A self-replicating hexadeoxynucleotide. *Angewandte Chemie, International Edition in English*, 25, 932–935.

97. von Kiedrowski, G. (1993). Minimal replicator theory I: Parabolic versus exponential growth (pp. 115–146). In *Bioorganic chemistry frontiers*, Vol. 3. Berlin: Springer-Verlag.
98. Wachterhauser, G. (1997). The origin of life and its methodological challenges. *Journal of Theoretical Biology*, 187, 483–494.
99. Walde, P., Wick, R., Fresta, M., Mangone, A., & Luisi, P. L. (1994). Autopoietic self-reproduction of fatty acid vesicles. *Journal of American Chemical Society*, 116, 11649–11654.
100. Watanabe, K., Ferrario, M., & Klein, M. L. (1988). Molecular dynamics study of a sodium octanoate micelle in aqueous solution. *Journal of Physical Chemistry*, 9, 819–821.
101. West, G. B., Brown, J. H., & Enquist, B. J. (1997). A general model for the origin of allometric scaling laws in biology. *Science*, 276, 122–126.
102. West, G. B., Woodruff, W. H., & Brown, J. H. (2002). Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals. *Proceedings of the National Academy of Sciences of the U.S.A.*, 99, Suppl. 1, 2473–2478.
103. Whitten, D., Chen, L., Geiger, H., Perlstein, J., & Song, X. (1998). Self-assembly of aromatic-functionalized amphiphiles: The role and consequences of aromatic-aromatic noncovalent interactions in building supramolecular aggregates and novel assemblies. *Journal of Physical Chemistry B*, 102, 10098–10111.
104. Wills, P., Kauffman, S., Stadler, B., & Stadler, P. (1998). Selection dynamics in autocatalytic systems: Template replicating through binary ligation. *Bulletin of Mathematical Biology*, 60, 1073–1098.
105. Woodruff, W. (2002). Private communication. This proto-organism work ties in with the very active area on universal scaling relations in biosystems, as it defines the origo in the charts used. A comprehensive discussion of this active interdisciplinary research area is being carried on in the LDRD-DR project “Scaling relationships in biology” at Los Alamos National Laboratory, 2002–2005 (PI: D. Breshears, EES Division).
106. Yoo, K. H., Ha, D. H., Lee, J. O., Park, J. W., Kim, J., Kim, J. J., Lee, H. Y., Kawai, T., & Choi, H. Y. (2001). Electrical conduction through poly(dA)-poly(dT) and poly(dG)-poly(dC) DNA molecules. *Physical Review Letters*, 87(19), 198102.